Open Journal of Legal Science 法学, 2024, 12(1), 231-241
Published Online January 2024 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2024.121034

Hans汉斯

生成式人工智能的个人数据风险及其法律规制

刘　迈

扬州大学法学院，江苏　扬州

收稿日期：2023年10月16日；录用日期：2023年10月25日；发布日期：2024年1月15日

摘　要

以ChatGPT为代表的新一代生成式人工智能技术，具备自主学习性、强交互性、创新性等特点，对既有的个人数据风险治理体系构成了挑战。为应对其引发的个人信息安全风险，有必要从生成式人工智能的数据收集与处理、算法运行与其生成性内容三个关键要素进行探讨。针对个人数据侵权、算法偏见以及生成性内容的非法利用等问题，需要进行分类讨论。结合我国已有的人工智能立法实践分析，其对生成式人工智能的个人信息规制存在法律规范缺位、规则适用困难、治理手段乏力等现实难题，需从立法、法律适用、审查监督、司法救济四个层面提出相应的法律规制建议，以促进生成式人工智能的个人数据风险法律规制体系更加严谨和完善。

关键词

生成式人工智能，ChatGPT，个人信息，风险治理

Personal Data Risks of Generative Artificial Intelligence and Its Legal Regulation

Mai Liu

School of Law, Yangzhou University, Yangzhou Jiangsu

Received: Oct. 16th, 2023; accepted: Oct. 25th, 2023; published: Jan. 15th, 2024

Abstract

The new generation of generative AI technologies represented by ChatGPT, characterized by autonomous learning, strong interactivity, and innovativeness, poses a challenge to the established personal data risk governance system. In order to deal with the risk of personal information security caused by it, it is necessary to explore the three key elements of generative AI, namely data collection and processing, algorithm operation and its generative content. The issues of personal

文章引用: 刘迈. 生成式人工智能的个人数据风险及其法律规制[J]. 法学, 2024, 12(1): 231-241.
DOI: 10.12677/ojls.2024.121034

data infringement, algorithmic bias, and illegal utilization of generative content need to be discussed in a categorized manner. Combined with the analysis of China's existing AI legislation and practice, the regulation of personal information of generative AI is characterized by the absence of legal norms, the difficulty of applying rules, and the lack of governance tools, etc. It is necessary to put forward the corresponding legal regulation proposals from the four levels of legislation, application of law, review and supervision, and judicial relief, so as to promote a more rigorous and perfect legal regulation system for the risk of personal data of generative AI.

Keywords

Generative Artificial Intelligence, ChatGPT, Personal Information, Risk Governance

1. 引言

自 2022 年 11 月以来，针对 OpenAI 公司发布的生成式人工智能聊天机器人 ChatGPT，学界对于新一代生成式人工智能的法律风险规制问题进行了广泛而深入的探讨。2023 年 3 月 24 日，OpenAI 公司指出，由于其开源数据库可能存在的错误，导致 ChatGPT 缓存出现问题，部分用户可能看到其他用户的历史聊天记录以及信用卡的最后四位数字、到期日期、姓名等个人信息。约 1.2%的用户个人信息可能受到此漏洞的影响。这一事件引发了学界对生成式人工智能可能引发的个人信息安全法律风险的关注。大多数学者表达了对 ChatGPT 用户数据隐私以及个人信息存在泄露风险的担忧[1]。与传统的大数据为基础形成预测结果的决策式人工智能不同，新一代生成式人工智能可以通过学习海量的人类创造的内容来生成新的内容[2]，其创造能力得到很大程度提升。然而，生成式人工智能的运行高度依赖于数据和算法的支持，对于像 ChatGPT 这类高风险的生成式人工智能，如何对其数据处理和算法运行进行合理规制是预防其个人信息法律风险的关键所在。本文从多个方面分析了生成式 AI 运用过程中可能引发的个人信息安全法律风险，并结合我国关于生成式人工智能的立法实践，探讨了新一代生成式人工智能个人信息安全风险规制的可选择途径。

2. 生成式人工智能技术导致的个人信息安全风险

目前，生成式人工智能主要包含两种类型：生成式对抗网络(GAN)和生成式预训练转化器(GPT) [3]。以 ChatGPT 为例，其主体架构可分为三个阶段：1) 语料数据收集阶段：ChatGPT 从各种资源中收集信息，并形成海量的文本数据基础。2) 预训练算法与模型实现预训练大规模语言模型：在具备充分的语料基础后，ChatGPT 被赋予理解自然语言和上下文生成自然语言的能力。3) 微调阶段：这个阶段是由 OpenAI 研发的 Codex 模型完成的，它赋予了 GPT 模型代码生成和代码理解的能力，使得它能够生成的答案更加合理。ChatGPT 具备的内容生成能力基于对大规模数据的收集与处理。这种能力会随着新数据的不断涌入而不断升级，但在提升能力的同时，也产生了巨大的个人信息安全隐患。

生成式人工智能所引发的个人信息风险，并非仅限于其数据处理过程中的某一特定环节或领域，而是贯穿于算法机制对数据的整个动态利用过程中，并取决于生成属性的具体表现。因此，探究生成式人工智能的个人信息安全风险来源，应围绕数据、算法和生成内容这三个核心要素进行深入分析。

2.1. 个人数据层面安全风险来源

在大数据背景下，个人数据已经成为个人信息最主要的载体[4]，本文将重点关注与个人信息密切相关的生成式人工智能在处理个人数据时可能产生的个人数据安全风险，这些风险对个人信息主体的个人信息财产权益和人格权益具有直接影响。

2.1.1. 个人数据来源合法性风险

《中华人民共和国个人信息保护法》(以下简称“《个人信息保护法》”)第十四条规定，个人信息处理者应当取得个人的明确同意。因此，在收集用户个人信息时，无论是普通互联网应用还是生成式人工智能应用，均应通过隐私政策或个人信息法律保护政策，告知用户其个人信息收集行为，并由个人信息主体自主决定是否允许个人信息被收集。另外需要注意的是，生成式人工智能涉及多阶段数据收集，可能会出现个人数据被混合收集的情况。因此，生成式人工智能的个人数据来源合法性具有较大的法律风险。

一方面，生成式人工智能在预训练阶段需要对大量数据进行收集。由于该阶段并未完全遵循“通知－同意”原则，导致知情同意原则在此阶段失去其约束力。另一方面，当生成式人工智能在运行阶段时，例如 ChatGPT，会收集用户账户信息、用户内容、通信信息、社交媒体等个人信息[5]，如果用户拒绝提供个人信息，则可能无法获得完整的服务。许多互联网应用程序需要用户在注册时同意隐私协议，否则无法使用该软件，这种现象实际上是一种对用户信息的不当收集[6]。在 AI 内容生成阶段，OpenAI 公司的一项隐私策略指出，用户与应用进行对话时所提出的问题以及应用生成的内容也将被自动收集。在此过程中，用户的个人信息可能会以生成内容的形式被再次收集，这种行为并未在使用过程中告知用户。

2.1.2. 个人数据非法使用风险

除了遵循“知情同意”原则外，《个人信息保护法》第 6 条规定，信息处理者应遵守“目的限制”原则，即要求企业或相关主体在收集个人信息时，必须具有“具体、清晰和正当的目的”。在处理个人信息时，不得违反初始目的[7]。因此，生成式人工智能在收集个人数据后，仍存在非法使用个人数据的可能性。

第一，存在个人数据泄露风险。由于生成式人工智能技术的错误，可能导致个人数据泄露。个人数据的存储措施是否合规是保障个人数据存储安全的关键。目前，OpenAI 公司未提供向用户个人提供检查其个人数据存储库的方式，且发布的信息使用条款并未详细说明用户个人信息的存储期限、具体保护措施和救济方式。根据《个人信息保护法》第 17 条第 2 款规定[1]，个人信息存储应有时间限制，一般为实现处理目的所必要的最短时间，并非可以无限期保留。但实践中，ChatGPT 这种对信息存储期限暧昧不明的做法，不利于保障用户的个人信息权益。

第二，存在个人数据被非法商业利用的风险。根据 OpenAI 公司的隐私策略中的公开条款，该公司在某些情况下会向第三方提供个人信息，其中包括用户的商业信息和网络活动信息，这些信息可能加剧人工智能决策的算法歧视问题。此外，该公司在某些情况下也会在用户不知情的情况下向第三方提供其个人信息。

第三，存在个人数据跨境流动风险。在生成式人工智能领域，如何恰当存储个人数据的问题引发了广泛质疑，同时，其带来的个人数据跨境流通风险也受到了广泛关注。数据作为国际竞争与合作的重要资源，个人数据不仅涉及公民的个体利益，更进一步涉及国家数据主权安全。首先，任何国家对他国公民的个人信息进行非法收集都可能会引发数据垄断与数字霸权等问题，这是对全球数字秩序的严重挑战。其次，个人信息的大范围传播很可能会引发国家情报安全问题。一旦大量的敏感个人信息被人工智能非

[1]《中华人民共和国个人信息保护法》第十七条第 2 款规定：“个人信息处理者在处理个人信息前，应当以显著方式、清晰易懂的语言真实、准确、完整地向个人告知个人信息的保存期限。”

法传输，就可能引发严重的“数据窃取”以及“数据攻击”等安全问题，对各国的国家安全构成了严峻的挑战。以 ChatGPT 为例，其用户注册的个人信息以及后续对应用提出的问题都会被传输到美国的 OpenAI 公司，以便其利用该数据与美国的第三方主体进行合作。若我国用户所提的问题涉及个人信息、敏感信息甚至涉及国家安全、公共健康和安全等方面的重要数据，则存在着极大的法律风险。

2.2. 算法运行层面安全风险来源

如前所述，如果在对生成式人工智能的数据收集和处理阶段未能实施必要的规范和限制，其负面影响将一直延续到算法运行阶段。人工智能算法的歧视问题一直存在，算法将人们在网络上的习惯、偏好、购物记录、GPS 位置数据等各种网络活动转化为对人们的各种评分和预测。在算法的影响下，对个人数据的个人信息进行掌握和分析是不可避免的，个人信息主体因此成为生成式人工智能所计算的客体。

2.2.1. 算法黑箱侵犯个人数据主体权利

随着生成式人工智能能力的显著提升，所带来的人工智能算法黑箱问题也日益复杂。算法黑箱，指的是人们无法直接观察或打开算法数据系统以了解其数据处理过程，只能获取输入数据和输出数据[8]。简单来说，当我们利用人工智能进行决策或生成性工作时，我们无法了解其决策的推论过程和生成性内容的数据来源。因此，在司法实践中，很难准确判断其生成性内容是否存在对个人信息的不当利用。算法黑箱容易导致生成式人工智能违背个人信息“知情同意”原则，并可能造成潜在的个人信息损害。

2.2.2. 数据偏见引发的算法偏见与算法歧视

人工智能的歧视性问题主要源于其背后的算法训练数据。在生成式人工智能算法中，由于使用了大量的个人数据进行训练和处理，因此可能会产生偏见或歧视。ChatGPT 是一种基于文本语料库和 RLHF 强化训练 [2] 的产物，它所表达的是文本数据、强化算法和系统设计者的价值观。尽管其文本输出看似中立和客观，但实际上仍然是算法和其背后操纵者的意识共同作用的结果。因此，由个人数据引发的算法偏见和歧视问题也需要引起重视[9]。

首先，从算法偏见的角度来看，ChatGPT 算法相较于传统算法模型，不仅依赖于自身的机器学习能力，还融入了许多人为因素[10]。然而，由于人工标注过程中可能存在的个人偏好以及机器学习算法框架本身可能存在的偏见，这种技术形式的负面效应被放大，导致算法偏见的产生渠道更加多样且难以预防。

其次，在算法歧视方面，例如在算法不当干预下产生的价格歧视、算法就业歧视等[11]。具体到生成式人工智能层面，由于个人数据和人为干预的影响，其内部算法歧视问题可能会外化成歧视性生成内容。这种歧视性生成内容可能导致对某些人群的不公平对待。因此，为了减少这种歧视问题，需要在算法设计和数据收集和处理过程中采取更加谨慎和公正的措施。

2.3. 生成性内容层面安全风险来源

相较于传统的决策式人工智能，新一代生成式人工智能的核心能力主要体现在其生成能力，这赋予了其一定的“创新”能力。以 OpenAI 公司最近发布的 ChatGPT-4 为例，它不仅能够依据存储和重复的知识进行推理和决策，还展示出比传统决策型人工智能更强的创造性和协作性。具体来说，GPT-4 不仅可以创作歌曲、编写剧本、生成各类满足用户需求的文本，其编程能力也得到了进一步提升。不过，生成式人工智能能力的提升也带来了一些个人信息安全风险。首先，由于生成式内容可能泄露个人信息。根据 OpenAI 公司的隐私策略，用户与应用进行对话过程中所提出的问题和应用生成的内容都将被自动收集作为一种数据，用户与应用对话的过程实际上就是其个人信息被收集的过程。因此，用户的个人信息可能会以生成内容的形式被泄露。

[2]RLHF 是一种基于强化学习的算法，通过结合人类专家的知识和经验来优化智能体的学习效果。

其次，GPT-4 的发布也降低了攻击代码编写的技术门槛，使得无代码编程成为可能，这导致个人信息可能被生成的恶意程序盗取的风险大大增加。因此，在享受新一代生成式人工智能带来的便利和新颖性的同时，我们也需要关注其可能带来的个人信息安全风险。

3. 生成式人工智能个人信息风险法律规制难题

在宏观层面，中国已初步建立起涵盖法律、部门规章、地方性法规的多层次人工智能治理规范结构。这一结构以《个人信息保护法》、《中华人民共和国网络安全法》(以下简称“《网络安全法》”)、《中华人民共和国数据安全法》(以下简称“《数据安全法》”)为核心，构成了中国的人工智能治理体系。2023 年 7 月 10 日，国信办公布了《生成式人工智能服务管理暂行办法》(以下简称“《办法》”)，以促进生成式 AI 的健康发展，并对相关应用进行规范。该《办法》共包含 21 条，主要明确了办法的适用范围、生成式人工智能的定义、责任主体的认定，还为生成式人工智能服务主体设立了数据合规责任。

虽然当今我国针对网络空间进行法律规制的基本制度框架已趋完成[12]，我国对于人工智能治理的立法实践正处于稳步推进阶段，在以 ChatGPT 为代表的生成式人工智能以及深度合成等领域，立法规范在全球范围内甚至处于领先地位。然而从实际的角度来看，我国目前在以个人信息保护为核心的生成式人工智能个人信息风险的规制方面仍然存在诸多挑战。

3.1. 规范构建缺位：人工智能个人信息保护法律规范仍未完善

我国目前在生成式人工智能个人信息风险法律规制方面面临的首要挑战是规范构建的缺失。从专业性角度来说，人工智能领域的立法在个人信息保护方面仍然缺乏一定的专业性和针对性，因此未能充分预防相关风险的发生。此外，从监管角度看，虽然当前的法律规范中设置了多个监管主体的措施，有利于规范各领域的人工智能法律风险，但是过多的监管主体参与到治理过程中也可能会带来诸如竞争、推诿等问题，影响了治理效果。最后，从完备性角度看，《办法》等治理规范尚不完备，存在法律漏洞，这也影响了个人信息风险的有效规制。

3.1.1. 生成式人工智能领域个人信息保护立法专业化程度欠缺

根据我国目前对人工智能立法的实践，可以观察到立法在推动人工智能发展的同时，并未忽视对个人信息和数据安全的保护[3]。然而，从已制定的规范来看，人工智能领域立法对于个人信息安全的保护条款相对较为笼统，缺乏专业性和针对性，因此尚未充分发挥风险预防的作用。深入分析《办法》的具体条款可以发现，仅在生成式人工智能法律中简单列入个人信息保护条款，实际上仍属于传统的事后规制立法模式。这种做法可能无法对个人信息安全风险形成有效的预防和管控。

从数据治理的角度来看，《办法》虽然规定了生成式人工智能服务提供主体对个人数据来源的合法性义务，但目前的数据治理状况难以实现互联网数据的清晰划分和追溯。因此，对生成式人工智能所收集的个人数据来源进行追溯并不现实。

另外，从生成性内容的治理来看，《办法》第 4 条规定利用生成式人工智能生成的内容应真实准确。然而，该条款并未明确解释“真实准确”的具体内涵，也没有明确规定生成式人工智能服务提供者的过错认定标准。这可能给实际操作带来一定的不确定性和难度。

3.1.2. 责任主体的复杂性与重合性导致监管难问题

深度合成技术是指利用深度学习、虚拟现实等技术来制作文本、图像、音频、视频、虚拟场景等信

[3]《生成式人工智能服务管理办法(征求意见稿)》第 4 条第 5 款规定，生成式人工智能的生成内容应该尊重他人合法利益，防止伤害他人身心健康，损害肖像权、名誉权和个人隐私，侵犯知识产权。禁止非法获取、披露、利用个人信息和隐私、商业秘密。《深圳经济特区人工智能产业促进条例》第 72 条第 2 款规定，开展人工智能研究和应用活动，不得侵犯个人隐私或者侵害个人信息权益。

息[13]，是生成式人工智能生成性内容的重要手段。由于生成式人工智能的实现需要数据、算法和深度合成技术的结合，因此生成式人工智能服务提供主体也可能同时是算法推荐服务提供者和深度合成技术服务提供者。在这种情况下，当生成式人工智能个人信息侵权行为发生时，就可能出现责任主体竞合和监管主体不明确的问题。

目前，我国对人工智能的监管呈现出多渠道、多部门的现状。多个部门，包括国家市场监督管理总局、国家互联网信息办公室、工业和信息化部、科技部等，都有参与人工智能的监管工作。这种多监管主体的设置有利于规制多领域的人工智能法律风险。然而，过多的主体参与到治理过程中也可能产生新的问题。例如，监管责任的竞合可能导致监管部门之间的竞争和推诿，这在一定程度上还会引发利益冲突，阻碍执法工作。面对复杂的情况和涉及广泛的范围，各个监管主体可能会选择避而不谈。

3.1.3. 生成式人工智能治理规范仍存在法律漏洞

当前人工智能治理规范尚未完善，仍存在法律漏洞。从《办法》的具体规定来看，虽然对生成式人工智能服务提供者的主体责任进行了规定，明确了生成式人工智能产品的技术研发商、应用开发商、提供 API 接口等接入服务的提供商均需要为其生成的内容承担责任，但并未对生成式人工智能服务的用户责任进行明确。由于生成式人工智能具有生成性，其风险不仅存在于生成式人工智能服务的内部，外部风险也需受到法律规制。OpenAI 发布的 GPT-4 降低了攻击代码编写的技术门槛，具备了无代码编程能力，在此过程中，个人信息也面临着被生成的恶意程序盗取的巨大风险。目前要求生成式智能服务提供者全流程审查监管是不可能的，因此有必要对生成式人工智能的用户责任进行补充规定，避免用户利用其服务侵犯他人个人信息合法权益的内容，明确用户责任与平台责任的界限，适当地减轻平台的风险保障义务负担。除此之外，在生成式人工智能个人数据去识别化与跨境流通的规制方面，仍未有相关立法进行明确。

3.2. 规则适用困境：个人信息保护规则适用困难

第二个规制难题是规则适用困境。从法律赋权的角度看，在生成式人工智能等新一代人工智能技术的冲击下，《个人信息保护法》赋予自然人的相关个人权利被逐步消解和虚化。从法律规定统一标准的角度看，由于生成式人工智能对个人数据的范畴和边界未明确，目的限制原则的标准难以统一界定。

3.2.1. 生成式人工智能对“知情同意”规则的消解

我国《个人信息保护法》通过赋予权利的方式，确立了自然人的个人信息权，保障个人信息主体在个人信息的处理过程中享有知情权、决定权[4]。然而，在大数据与人工智能技术的冲击下，以个体控制权为核心的知情同意规则逐渐失灵，导致我国个人信息保护法的实施效果逐渐减弱。

首先，从信息主体层面来看，个人信息控制权是建立在个体对个人信息理性支配的前提下的，但在实践中，个体对人工智能信息收集存在理解上的壁垒，无法真正了解人工智能对个人信息的收集范围和利用程度。因此，作为信息收集一方的人工智能架空了个体的权利，导致个人无法支配自身的个人信息。因此，大多数人也已经接受了“数据裸奔”的状态，其知情同意权也成了一种可有可无的权利。

其次，就以 ChatGPT 为代表的生成式人工智能而言，其隐私政策将收集用户个人信息作为一种理所当然的权利，但其处理信息的范围、目的方式、保存期限、算法推荐方法等都未进行有效通知，导致用户无法完全知晓个人信息被收集后的用途，也不存在真正意义上的同意。因此，生成式人工智能直接削弱了或剥夺个人信息的自决权。

另外，在预训练阶段，生成式人工智能便对个人数据进行了收集，而在此过程中我们无法期待生成

[4]《中华人民共和国个人信息保护法》第四十四条规定：“人对其个人信息的处理享有知情权、决定权，有权限制或者拒绝他人对其个人信息进行处理；法律、行政法规另有规定的除外。”

式人工智能服务提供者履行其个人数据收集通知义务。这也进一步削弱了个人信息保护的效果。

3.2.2. 生成式人工智能个人数据处理的目的限制原则标准难统一

我国《个人信息保护法》第 6 条[5]明确规定了目的限制原则。根据该原则，个人信息处理者在处理个人信息前，必须明确其处理该信息的目的，且该目的必须是适当、相关和必要的。个人信息处理者的处理行为不得超出信息主体初始授权的范围。然而，对于生成式人工智能中个人数据的范围及边界，目前尚未有明确的定义。举例来说，OpenAI 公司的隐私策略中的公开条款指出，在某些情况下，除非法律要求，否则该公司在用户不知情的情况下可能会将其个人信息提供给第三方，包括供应商和服务提供商等。这其中包括用户的商业信息和网络活动信息。在这种情况下，目的限制原则的贯彻执行可能面临一定的困难。

3.3. 治理手段局限：司法救济缺位

第三个规制难题是治理手段的局限性。从相关法律框架来看，我国主要依靠行政监管来规范生成式人工智能的发展，但在人工智能和大数据技术的冲击下，个人信息受到损害的特性，如无形性、潜伏性、未知性以及难以评估等，变得更加突出[14]。此外，像 ChatGPT 这样的生成式人工智能也对侵权责任制度产生了深远影响。这类人工智能造成的侵权行为具有复杂的侵权主体、智能化的侵权行为以及多元化的因果关系等特点[15]。

“举证难”一直是个人信息主体维权的一个重要障碍。尽管现有的个人信息侵权责任判定采用过错推定责任原则，这在一定程度上减轻了数据弱势群体的举证责任，并强化了相关信息处理平台的注意义务。但是，由于生成式人工智能对个人信息造成的损害往往不易被察觉，这就进一步加剧了“举证难”的问题。在这种情况下，确定生成式人工智能造成的个人信息损害标准，将司法作为个人信息风险治理的一种最终保障方式，就变得尤为重要。

除此之外，由于人工智能本身的复杂性和专业性，对其治理需要体现多主体的共同参与。因此，有必要构建一个全面而有效的多元主体共同治理的生成式人工智能治理模式。

4. 生成式人工智能个人信息风险法律规制策略

为针对生成式人工智能所带来的个人信息风险构建合理的法律规制策略，我国应结合国内外已有的生成式人工智能立法实践，以及当前我国生成式人工智能个人信息规制存在的法律规范缺位、规则适用困难、治理手段局限等现实难题，从立法、法律适用、审查监督、司法救济四个层面提出相应的法律规制建议，以促进生成式人工智能个人信息法律治理体系的完善。

4.1. 规范构建：完善多维度人工智能个人信息法律保障体系

为应对新一代生成式人工智能发展带来的个人信息安全伴生风险，保障公民基本权利的确定性，国家应尽义务事先制定法律以预防可能的未来风险。因此，完善生成式人工智能个人信息法律保障体系，强化风险预防在立法中的体现，仍是我国当前的主要任务。尽管《办法》在一定程度上体现了风险预防的理念，但与欧盟《人工智能法案》相比，其发挥风险预防作用仍有不足。因此，在后续的生成式人工智能立法中，应继续强化法律的风险预防作用。

4.1.1. 个人数据来源合法化治理

为确保新一代生成式人工智能的合法性与合规性，应制定相关法律法规，明确数据收集的来源与处

[5]《中华人民共和国个人信息保护法》第六条规定：“处理个人信息应当具有明确、合理的目的，并应当与处理目的直接相关，采取对个人权益影响最小的方式。收集个人信息，应当限于实现处理目的的最小范围，不得过度收集个人信息。”

理方式。第一，针对不同来源的个人数据，应采取不同的处理方法。1) 如果生成式人工智能应用直接收集了公民的个人信息，那么这种收集行为必须严格遵守《个人信息保护法》的相关规定。在收集之前，必须征得个人信息主体的明确同意，同时还要明确规定数据的存储期限和利用目的。在后续的个人数据利用过程中，仍需遵守目的限制原则。2) 如果生成式人工智能应用所使用的个人信息来自已经公开的数据，那么虽然根据《个人信息保护法》第 27 条的规定，开发者可以对此类个人信息进行处理，但对此类个人信息的利用仍需遵守目的限制原则。3) 应禁止生成式人工智能应用通过爬虫等非法手段获取数据，确保此类个人数据的收集不会脱离法律的监管。

第二，我们应当进一步细化生成式人工智能在个人数据收集、处理环节中的规范流程，改变目前个人数据与其他数据在管理上混淆不清的状况。生成式人工智能的应用不仅需要个人数据作为支撑，同时还需要从互联网中获取大量的其他类型数据来进行训练和优化。因此，我们在进行数据收集和处理时，必须对个人数据的来源进行严格的审查，并采取更加严密的保密措施和存储方案来确保所收集的个人信息安全可靠并得到妥善保存。

第三，针对不同类别的个人信息应采取相应的法律保护策略。首先，立法应加强对敏感个人信息的法律保护，并审慎评估生成式人工智能在这些领域可能带来的风险。例如，在医疗健康、人脸识别、生物基因检测等领域，应禁止生成式人工智能收集和利用这些个人信息，并设置更为严格的备案制度和准入门槛，以避免对个人信息主体造成不可挽回的损害。其次，应促进生成式人工智能个人数据的去识别化管理。虽然个人信息和个人数据之间存在一定的关联，但它们并不完全等同。因此，个人数据应分为两类进行讨论：① 已经去识别化的个人数据；② 未去识别化或去识别化后仍具有识别性的个人数据。生成式人工智能所收集数据的客观性和准确性不仅会影响其算法的透明度和精准度，还会对其生成性产品的内容产生影响。因此，生成式人工智能服务提供者应进行个人数据的匿名化处理，这不仅有利于保障个人信息主体的个人数据权益，还有利于促进生成式人工智能数据的个人数据治理，并明确区分个人信息权和企业个人数据所有权之间的不同权利话语体系[4]。

4.1.2. 优化责任承担及监管主体结构

在规定责任主体方面，仅对生成式人工智能服务提供者的责任进行规范是不足够的，使用者不应被排除在侵权责任承担主体之外。首先，通过完善立法来规定生成式人工智能使用者的责任，有利于打击恶意利用该人工智能技术盗取个人信息以及破坏数据安全生态的行为。将平台责任与个人责任分开，能够减轻生成式人工智能服务提供主体的负担，并形成由服务提供者、技术开发者及服务使用者构成的三方权责机制，从而实现生成式人工智能风险责任的合理、有效分配。其次，应当简化生成式人工智能的监管主体，明确监管责任主体为生成式人工智能的数据、算法和生成性内容。构建多层级、分阶段的监管体系，各司其职并在监管过程中相互配合，形成监管合力。

4.1.3. 生成式人工智能个人数据跨境流通的漏洞弥补

随着大数据和人工智能技术的不断发展，个人数据跨境流通引发的国家数据安全风险日益凸显。学术界已将网络空间纳入国际法对“境”的定义范畴，但对于生成式人工智能个人数据的跨境问题，仍需立法予以进一步明确。

首先，应完善生成式人工智能数据出境评估机制。个人数据的跨境流通应经过相关组织和监督机关的安全评估，并对不同性质的数据采取差异化的管理评估方法，以严格保障敏感个人信息的跨境流动安全。2022 年 5 月 19 日，国家互联网信息办公室通过了《数据出境安全评估办法》，该办法对出境数据的类型、数量及评估办法等方面作了规定。个人数据出境需经过初步安全评估，并针对涉及国家数据安全、个人数据权利保护等不同风险类型的数据出境采取不同的管控措施。

此外，除了合理设定各方权利义务、满足相应的标准和程序外，个人数据出境还需获得原数据主体的单独同意和明确授权，并适当强化监管，建立多阶段式的个人数据跨境流动监管机制。

4.2. 规则适用：知情同意与目的限制原则的完善

为应对生成式人工智能法律规制难题之一的规则适用困境，我国相关部门应健全生成式人工智能个人信息收集、处理通知规则，以深化知情同意规则体系。此外，相关部门还需尽可能地限制生成式人工智能服务提供者个人数据的使用目的，以实现目的限制原则标准的统一。

4.2.1. 深化知情同意规则体系

首先，有必要建立健全的生成式人工智能在个人信息收集、处理环节的通知规则。除了在收集用户个人信息时应当进行通知外，还应在后续的个人数据利用以及人工智能自动化处理的情形下，数据控制者除应向用户披露一般的信息之外，还应特别向用户披露包括其自动化处理的逻辑程序信息、处理该信息的重要性以及其对于用户可能造成的影响，增加了数据控制者的风险通知义务。同时，数据控制者应以简洁、清晰且易于理解的方式进行通知，确保通知的有效传达。

其次，完善同意机制是必要的。无论数据控制者通过何种方式对用户的个人信息进行利用，都应当取得用户的明示同意。同时，还应依据个人信息的敏感程度区分个人信息的类型，对于个人敏感信息的处理设置更加严格的同意要件。例如，对于涉及儿童个人信息的收集及处理，应当特别取得其监护人的同意[16]。

4.2.2. 统一目的限制原则标准

由于生成式人工智能在使用个人数据时的目的和程度存在差异，因此制定一套适用于各种情况的通用范式规范是极其困难的。未来，生成式人工智能将会广泛应用于现代社会的各个领域，这些领域涉及的数据种类和敏感程度各不相同。考虑到前文所述，应当建立针对不同场景的生成式人工智能个人数据利用标准，对生成式人工智能服务提供者使用个人数据的目的进行限制。在变更使用目的之前，应事先征得知情同意原则的前提条件，以避免个人信息完全脱离信息主体的控制范围。

4.3. 审查监管：平台审查与行政监管相结合

当今我国针对网络空间进行法律规制的基本制度框架已趋完成，而网络治理的关键主体和核心议题一直以来都是网络服务提供者和网络服务提供平台责任[17]。相关部门应当对生成式人工智能服务提供平台的隐私政策进行审查，对以 ChatGPT 为代表的算法模型在投入国内正式投入应用之前进行严格的法律审查，以达成平台审查与行政监管相结合的综合治理手段。

4.3.1. 统一隐私政策规范标准

一方面，我们强烈建议推动生成式人工智能企业成立行业协会，并发布统一的个人数据收集与处理行业技术指南，以及制定生成式人工智能个人数据处理的基础性标准。例如，推特公司发布的世界首个反对“深度伪造”方案、谷歌公司发布的 AI 原则等。通过行业规范的建立和完善，我们可以更好地完善生成式人工智能应用的个人信息保护机制，并建立可信赖的人工智能体系，以更合理的方式利用个人数据[18]。

另一方面，相关监管部门应对生成式人工智能的隐私政策进行审查，以确保其隐私政策和个人信息使用条款的合法性和合规性。隐私条款本质上是一种格式条款，由于条款提供方通常具有信息优势和缔约地位，往往存在明显的利益偏向性[19]。因此，预先确定隐私政策的标准，并嵌入个人信息权益保障条款，可以有效解决“知情同意”的虚化现状，从而保障用户对其个人信息利用的知情权不受侵害。

4.3.2. 审查监管促进算法纠偏

《算法推荐管理规定》第 8 条规定，算法推荐服务提供者有义务对其算法模型进行审查。在遵循该规范文件的前提下，ChatGPT 的算法模型在应用前应接受严格的法律审查，以避免机器学习过程中引入人为的算法偏见。同时，应将规范文件的要求转化为技术标准，并将其融入算法程序的编译过程中，以预防潜在的法律风险。

鉴于 ChatGPT 的特殊技术特性，对其进行法规管制可分为两个阶段。

第一阶段，针对生成式人工智能存在的先天性算法偏见，在算法程序编译环节进行预防。机器学习过程是将部分数据作为输入并产生相应的结论作为输出。其中，计算过程需要进行预先训练，这是算法的机器学习过程，应将规范文件的要求融入算法程序的设计过程中。在算法设计过程中，对于可能存在算法偏见的参数，应及早发现并消除，调整和校对算法程序的偏见，使其回归正常的算法运行路径，通过规范文件的约束来规范算法技术，避免算法偏见的持续放大。这一阶段旨在通过“技术治理”的工具赋能路径，完善对算法程序代码的监督样态。

第二阶段，加强对人工标注的算法偏见的审查与监管。人工标注行为可能带有强烈的个人性偏向。为解决这一问题，平台应预先设定人工标注的标准，并及时纠正明显带有算法偏见的标注。

4.4. 法律救济：强化生成式人工智能个人信息侵权司法救济

事后救济作为生成式人工智能风险预防的重要组成部分，发挥着司法救济在个人信息风险治理中的关键作用。这种安排有利于形成“平台审查－行政监管－司法救济”的多元主体治理环节，具体措施如下。

第一，建立健全生成式人工智能侵权责任追究机制。当前，中国立法尚未针对生成式人工智能侵权作出专门规定，因此在确定侵权责任时，需参照现有侵权理论进行责任归属。关于生成式人工智能是否能够成为独立的侵权法律主体，学界尚存在争议。有学者主张，在发生侵权行为时，应透过“人工智能的面纱”寻找真正的侵权责任承担主体[20]，在涉及个人信息侵权的情况下，常见的侵权主体包括生成式人工智能服务的提供者、用户和技术开发者。

在确定侵权归责原则时，应充分考虑生成式人工智能责任主体的复杂性和多元性。一般来说，对于一般侵权行为，应依据我国《民法典》中关于侵权责任的规定，采用过错责任原则进行归责。而在个人信息侵权方面，考虑到个人信息主体与生成式人工智能主体之间的差异，可采取无过错责任原则以减轻个人信息主体的举证责任。但同时，在具体实践中还要考虑生成式人工智能平台是否已履行了对个人信息的必要安全保护义务，以更好地保障数据主体的权益。

第二，发挥个人信息公益诉讼机制的作用。为了解决公民个人作为弱势群体在维护自身数据权过程中的困境，今后可以考虑引入集体诉讼机制。例如，可由生成式人工智能的行政监管责任主体或行业协会作为诉讼代表，针对不当收集和使用个人数据的生成式人工智能企业提起诉讼，从而弥补个人数据权主体在生成式人工智能数据治理中的劣势，切实保障个人信息主体的合法权益。

5. 结语

生成式人工智能对法律的挑战，看似难以应对，但却是当今社会必须承担的时代重任。中国共产党第十八届五中全会、第十九次全国代表大会、第十九届二中全会、三中全会和第二十次全国代表大会，先后提出了“网络强国”、“数字中国”、“智慧社会”和“网络生态”等战略部署。在第二十次全国代表大会的报告中，多处涉及互联网内容，既总结了历史性成就，又擘画了网络强国新时代的宏伟蓝图。司法部也明确提出了“数字法治”的建设要求。由此可见，“网络强国”、“数字中国”、“智慧社会”

和“网络生态”已成为党中央和我国政府的核心战略决策。在数字社会中，如何建设数字法治以及如何保障个人合法行使数字权利，已成为数字中国和智慧社会建设的重要议题。

在数字法治建设过程中，保障个人合法数字权益将是一项至关重要的任务。如前所述，生成式人工智能技术给个人信息安全带来了明显的风险。这种技术所带来的风险具有多层面、多阶段的特点，从预训练阶段的个人数据收集到算法运行再到最后生成性内容输出阶段，都可能对个人信息主体的合法权益产生不同程度的影响。因此，必须通过构建分阶段的个人信息风险规制体系，才能确保生成式人工智能的合规运行。以 ChatGPT 为代表的生成式人工智能是人类数字化社会发展的重要成就，在提供便利的同时，其伴生风险也引发了科技信任危机。法律只是预防其风险的一个组成部分，更深层次的问题是人与科技发展的平衡问题。对此，我们仍需要从法律、伦理等多个维度对其进行审慎考量。

基金项目

江苏省研究生科研创新计划：“新一代人工智能技术(ChatGPT)的法律风险及其宪法规制研究”(KYCX23_3479)。

参考文献

[1] 周智博. ChatGPT 模型引入我国数字政府建设: 功能、风险及其规制[J]. 山东大学学报(哲学社会科学版), 2023(3): 144-154.

[2] 於兴中, 郑戈, 丁晓东. 生成式人工智能与法律的六大议题: 以 ChatGPT 为例[J]. 中国法律评论, 2023, 50(2): 1-20.

[3] 毕文轩. 生成式人工智能的风险规制困境及其化解: 以 ChatGPT 的规制为视角[J]. 比较法研究, 2023(3): 155-172.

[4] 刘练军. 个人信息与个人数据辨析[J]. 求索, 2022(5): 151-159.

[5] OpenAI (2023) Privacy Policy. https://openai.com/policies/privacy-policy

[6] 郭雪慧. 人工智能时代的个人信息安全挑战与应对[J]. 浙江大学学报(人文社会科学版), 2021, 51(5): 157-169.

[7] 丁晓东. 大数据与人工智能时代的个人信息立法——论新科技对信息隐私的挑战[J]. 北京航空航天大学学报(社会科学版), 2020, 33(3): 8-16, 71.

[8] Bathaee, Y. (2018) The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*, **31**, 889-905.

[9] 邹开亮, 刘祖兵. ChatGPT 的伦理风险与中国因应制度安排[J]. 海南大学学报(人文社会科学版), 2023, 41(4): 74-84.

[10] 刘艳红. 生成式人工智能的三大安全风险及法律规制——以 ChatGPT 为例[J]. 东方法学, 2023(4): 29-43.

[11] 杨成越, 罗先觉. 算法歧视的综合治理初探[J]. 科学与社会, 2018, 8(4): 1-12, 64.

[12] 张欣. 从算法危机到算法信任: 算法治理的多元方案和本土化路径[J]. 华东政法大学学报, 2019, 22(6): 17-30.

[13] 张凌寒. 深度合成治理的逻辑更新与体系迭代——ChatGPT 等生成型人工智能治理的中国路径[J]. 法律科学(西北政法大学学报), 2023, 41(3): 38-51.

[14] 田野. 风险作为损害: 大数据时代侵权“损害”概念的革新[J]. 政治与法律, 2021(10): 25-39.

[15] 徐伟. 论生成式人工智能服务提供者的法律地位及其责任——以 ChatGPT 为例[J]. 法律科学(西北政法大学学报), 2023, 41(4): 69-80.

[16] 郑志峰. 人工智能时代的隐私保护[J]. 法律科学(西北政法大学学报), 2019, 37(2): 51-60.

[17] 张凌寒. 平台“穿透式监管”的理据及限度[J]. 法律科学(西北政法大学学报), 2022, 40(1): 106-114.

[18] 万志前, 陈晨. 深度合成技术应用的法律风险与协同规制[J]. 科技与法律(中英文), 2021(5): 85-92.

[19] 王俐智. 隐私政策“知情同意困境”的反思与出路[J]. 法制与社会发展, 2023, 29(2): 210-224.

[20] 袁曾. 生成式人工智能的责任能力研究[J]. 东方法学, 2023(3): 18-33.