Open Journal of Legal Science 法学, 2026, 14(9), 9-15
Published Online September 2026 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2026.149249

Hans汉斯

生成式人工智能应用中个人信息保护的法律困境与规范路径

喻扬晶

山东师范大学法学院，山东 济南

收稿日期：2026年7月22日；录用日期：2026年8月5日；发布日期：2026年9月1日

摘　要

随着生成式人工智能的广泛普及，个人信息处理的模式发生了根本性的转变，并对现行的个人信息保护法规体系提出了新的挑战。《个人信息保护法》中现行的相关规定，虽旨在保护个人信息安全，但其主要针对的是传统互联网环境，在应对生成式人工智能领域时，暴露出诸多适用难题。本文针对生成式人工智能在个人信息处理方面的特性，对现行法律在告知同意规则、目的限制与最小必要原则以及已公开个人信息处理等方面存在的适用困境进行分析。在此基础上，围绕告知同意规则、个人信息处理合法性基础以及服务提供者合规义务三个方面提出完善建议，以回应生成式人工智能应用中的个人信息保护困境，为相关制度的完善提供参考。

关键词

生成式人工智能，个人信息保护，法律适用，制度构建

Legal Dilemmas and Normative Pathways for Personal Information Protection in the Application of Generative Artificial Intelligence

Yangjing Yu

School of Law, Shandong Normal University, Jinan Shandong

Received: July 22, 2026; accepted: August 5, 2026; published: September 1, 2026

Abstract

The widespread application of generative artificial intelligence has transformed traditional modes

文章引用: 喻扬晶. 生成式人工智能应用中个人信息保护的法律困境与规范路径[J]. 法学, 2026, 14(9): 9-15.
DOI: 10.12677/ojls.2026.149249

of personal information processing and posed new challenges to the existing personal information protection regime. The current information protection rules under the *Personal Information Protection Law of the People's Republic of China*, which were primarily designed for traditional internet scenarios, face considerable difficulties when applied in the context of generative artificial intelligence. This article examines the characteristics of personal information processing in generative artificial intelligence and analyzes the practical dilemmas arising from the application of existing rules, particularly with respect to the informed consent rule, the principles of purpose limitation and data minimization, and the handling of publicly available personal information. On this basis, it proposes improvements in three key areas: the informed consent rule, the legal bases for personal information processing, and the compliance obligations of service providers, in order to provide a normative response to the challenges identified.

Keywords

Generative Artificial Intelligence, Personal Information Protection, Legal Application, Institutional Construction

1. 引言

生成式人工智能的迅猛发展所带来的一大变化是个人信息的处理模式得到了深刻变革，个人信息处理由传统的“小数据、特定目的、可预测”单向推进逐步向“大规模、动态化学习、持续生成”的新型处理模式发展。这一变革大大拓宽了传统个人信息处理模式的活动边界，个人信息被更为广泛且深度地进行运用，从而导致《中华人民共和国个人信息保护法》[1](后文简称《个人信息保护法》)中的多项制度规则面临着适用困境，同时也对现有的个人信息保护制度提出了新的规范挑战。

生成式人工智能对个人信息保护制度带来的挑战已引起学界广泛关注，现有研究大致可分为三类。第一类研究聚焦于技术特征与制度困境的分析。叶雄彪指出，既有个人信息保护规则在生成式人工智能背景下难以有效适用，知情同意权、查阅权、删除权面临被架空的困境[1]。武腾则目的限制原则在生成式人工智能应用中丧失实效，将连带影响最小必要原则等多重规则的适用[2]。第二类研究致力于保护模式的范式转换。付新华指出，需要构建“群体信息–个体信息”的双层保护体系[3]。李川则从刑法学视角提出，个人信息保护应从有限赋权模式转向场景治理模式[4]。第三类研究侧重于具体制度的调适方案，王若冰针对已公开个人信息提出“弱保护”的解释方案[5]。王苑从比较法视角考察了欧盟“正当利益”条款在 AI 预训练场景中的适用可能[6]。本文试图在上述研究基础上，以生成式人工智能的三个核心数据处理特征，即规模化、自动化、黑箱化为分析起点，探讨《个人信息保护法》的核心规则适用困境，并提出相应的制度完善方案。

2. 生成式人工智能对个人信息的数据处理特征

(一) 数据处理的规模化

生成式人工智能在模型升级与生成能力的提升过程中需要进行大规模的数据训练，而数据训练的频繁进行需要有海量的数据支撑作为基础。与主体、场景及目的都具有较强特定性的传统个人信息处理不

[1]https://www.spp.gov.cn/zdgz/202108/t20210820_527240.shtml

同，生成式人工智能对于语言、图像等内容的理解与生成能力的提升，需要收集大量的互联网公开数据、用户使用过程中的交互数据以及其他的多元化数据资源。在这一过程中，个人信息不再作为独立的数据被处理，而是被融入进了一个巨大的数据海洋之中集合处理，与其他种类的数据一同加入到模型训练和算法优化的过程中。数据规模的不断扩大导致个人信息的识别与使用边界逐渐模糊化，个人信息的主体往往对于自身信息是否被纳入模型训练以及被利用的具体方式并不具有完全的知情权与判断权。与此同时，数据处理的规模化也增加了个人信息被过度收集、超过授权范围的范围利用以及二次处理的风险，即使个人信息在处理过程中已经进行了匿名化或去标识化的处理，也可能因多个渠道的数据关联分析而被重新识别出特定的自然人，从而削弱现有个人信息保护制度的实际效果，传统的以单项个人信息为保护对象的个人信息保护制度因此面临新的挑战。

(二) 数据处理的自动化

生成式人工智能在数据处理过程还呈现出了高度的自动化特征。原始数据的收集与加工、算法模型升级以及最终内容物的生成等各个环节都是由算法自主完成[1]，人工干预在其中的占比非常低。在人为设置了某一特定目标后，算法会不断优化模型，同时在接收新的数据后再次进行动态化的更新，个人信息处理活动在这一过程中呈现出连续化、自主化的发展特征。在这一基础上，个人信息处理已不再针对某一具体主体、独立的处理行为，而是融入进了模型训练和算法优化的过程中，信息处理行为有着较强的动态性和隐蔽性。由于信息的处理活动是由算法自动完成，传统法律制度中以处理者的主观意思、具体处理行为为基础所建立的责任认定条款在适用上变得更加困难，个人信息处理责任主体的认定与归责更加复杂。此外，高度自动化的数据处理还增加了个人信息泄露、输出错误个人信息以及决策结果偏离等风险的出现，且一旦模型训练数据存在污染、算法存在缺陷，相关风险还可能随着模型输出而被进一步地放大与传播。

(三) 数据处理的黑箱化

生成式人工智能普遍采用的是深度学习等复杂的算法模型，其内部运行机制具有较强的不透明性，呈现出典型的“算法黑箱”特征。算法模型通过网络来对大量的数据进行学习和推理，在数据的输入学习、参数的计算以及内容生成物的输出之间缺乏一个明晰的处理过程，即使是模型开发者自身也难以准确地解释某一内容生成物产生的具体原因。对于个人信息而言，其在模型训练与使用中发挥了何种作用、是否对最终的生成内容产生了影响以及具体经由了何种处理路径，都无法追溯出一个特定明确的结果。这种黑箱化的特征导致信息在处理过程缺乏了必要的透明度，不仅增加了个人信息泄露、误用以及二次利用的风险，也使信息处理者法律规定的告知说明义务无法进行充分地履行，个人信息保护制度赋予信息主体的知情权、更正权、删除权等合法权利同样受到了较大的限制，无法得到正常实现。

3. 个人信息保护核心规则在 AI 场景中的适用困境

(一) 告知同意规则的适用困境

告知同意原则是判断个人信息处理行为是否具有合法性的关键原则，依据法律法规，信息处理者须对信息主体实施充分告知，并依法获取其同意，才能实施个人信息的处理行为。这一充分告知行为是信息主体能够有效作出同意决定的基础条件，然而，在生成式人工智能技术的具体应用领域，个人信息往往会经过数据收集、模型训练、模型更新优化等多个环节，不同阶段的数据用途还可能随着模型更新不断发生变化。在这种情况下，平台很难在信息处理开始之前就完整地向用户说明个人信息未来可能涉及的所有处理方式和使用场景。目前，大多数生成式人工智能平台主要通过“用户协议”或“隐私政策”履行告知义务，但相关内容往往较为原则，仅对数据收集、存储等事项作出概括说明，对于用户输入内容是否参与模型训练、训练数据如何管理、模型如何利用相关数据等问题缺乏具体说明。信息主体即使

阅读了相关协议，也难以真正了解自己所提供的个人信息将如何被处理，告知义务容易流于形式，难以达到保障用户知情权的目的[7]。

告知的目的在于保障个人信息主体能够在充分了解相关情况后自主决定是否提供信息，但在生成式人工智能的使用背景下，用户很难做出完全真实的意思表示。用户在使用生成式人工智能产品时，其提供给人工智能的文本、图片等内容可能被平台保存，并进一步用于模型训练或优化。虽然部分平台会在隐私政策中作出说明，但多数用户并不会仔细阅读相关条款，也难以理解复杂的算法规则，此时用户所谓的“同意”更多只是一种形式上的授权，但并不代表用户充分意识到了生成式人工智能会如何运用处理自己所提供的个人信息。此外，生成式人工智能从互联网上收集大量的公开数据来作为训练的材料，在这种情况下，信息的处理者与信息主体之间并未直接产生交流[8]，信息主体甚至不知道自己的个人信息已经被用于模型的训练，自然更谈不上作出是否同意的选择。传统告知同意制度需要建立在双方能够沟通的基础之上，而在大规模的数据抓取背景下，这一制度基础已经难以成立。

(二) 目的限制原则与最小必要原则的适用困境

《个人信息保护法》第六条确立了目的限制原则和最小必要原则，这两条原则将个人信息的处理活动的活动目的限制在明确、合理使用的范围内，同时要求收集个人信息时不得超过实现信息处理目的的必要限度。目的限制原则的确立目的在于限制个人信息的随意使用，也是信息主体判断是否同意个人信息处理的重要依据，然而，生成式人工智能的模型训练并不完全符合这一制度设定。大模型在最初的训练阶段需要从大量数据中学习语言规律和知识，其目的是形成具有通用能力的基础模型，而不是完成某一项具体任务。模型训练完成后，还可能根据不同的应用场景不断进行调整，信息处理的目的也会随着模型更新而发生变化。在模型训练开始时，信息处理者往往难以准确地向用户说明未来每一项数据处理活动的具体用途，也无法向用户列举模型后续可能涉及的全部应用场景。这样一来，传统意义上要求处理目的“明确、具体”的制度要求，在生成式人工智能使用过程中便难以得到完全的落实。

最小必要原则的实施前提是处理目的已经明确，在明确目的后才可以判断哪些信息属于“必要”，哪些信息属于“非必要”[2]。但对于生成式人工智能而言，模型的能力高度依赖于训练数据，模型训练不仅需要大量数据，还需要数据类型尽可能丰富，从技术角度来看，数据规模越大、数据质量越高，模型生成内容通常越准确、越稳定。因此，平台通常倾向于持续扩大训练数据规模，而不是仅收集完成某一具体任务所必需的数据。这意味着，最小必要原则强调的数据“少收集、少使用”与生成式人工智能模型升级所依赖的大量数据存在一定冲突，一些原本并非提供当前服务所必需的个人信息，可能会因模型训练和后续优化需要而被继续保存和使用，从而增加个人信息被长期保存、重复利用、甚至超范围使用的风险，也使最小必要原则在实践中的判断标准变得更加模糊。

(三) 个人信息质量保证义务的适用困境

《个人信息保护法》第八条规定，信息处理主体在处理个人信息应当保证个人信息的质量，避免因个人信息不准确、不完整对个人权益造成不利影响。而生成式人工智能的“生成”属性使得其在输出环节产生了传统数据处理中不存在的特有风险——“数据幻觉”。所谓数据幻觉，是指生成式人工智能在缺乏准确训练数据或逻辑推理支撑的情况下，生成看似合理但实际错误的信息。生成式人工智能在遇到生成数据不足的情况时，可能生成与事实不符的内容，其中包含虚构的个人信息，如不存在的姓名、职业经历、联系方式、信用记录甚至犯罪记录。这类信息既非来源于数据收集阶段对真实个人信息的处理，也难以追溯其具体的生成路径，而是模型在文本生成过程中自主“创造”的产物，

因此，对于生成式人工智能所造成的“数据幻觉”，其生成的“个人信息”并非来源于“处理”环节的不准确，而是 AI“生成”环节的产物，AI 生成错误信息的行为是否属于“处理”、是否应受第八条调整，谁是该信息的“处理者”，谁对生成内容的准确性负责，现行的《个人信息保护法》并未对其给出明

确的规定。即便认定第八条可以适用，“不准确”的判断标准也难以确定，AI 生成的内容究竟是“错误”还是“虚构”，在技术层面和法律层面往往难以形成统一的认识。且数据幻觉可能导致对个人名誉权、信用权等人格权益的侵害。与已公开个人信息被不当处理不同，此种侵害并非基于真实的个人信息，而是基于 AI 生成的虚假信息，其侵权形态更接近于诽谤或名誉侵权，但责任主体的认定因算法黑箱的存在而变得更加复杂。

(四) 已公开个人信息处理的合法性边界

《个人信息保护法》第二十七条规定，在遵循合理界限的前提下，信息处理主体可以对个人主动披露或已合法公开的个人信息进行处理，除非个人明确表示反对或该处理行为对个人权益构成显著损害。该规定旨在兼顾个人信息的流通利用与个人的合法权益保护，但并未进一步明确何种情形属于合理范围，也没有建立统一的判断标准。

在传统互联网场景下，已公开个人信息通常围绕着特定目的被进行有限地利用，例如搜索展示、新闻报道或者信息查询等，处理目的相对来说较为明确，处理的范围也较容易判断。生成式人工智能则有所不同，其对个人信息的利用方式明显超出了传统意义上的公开展示或者信息检索，但这种利用方式是否超出了合理使用的范围，法律并未进行明确的规定。如果将模型训练理解为已公开个人信息的合理利用，可能导致公开信息被无限度地大幅利益，不利于保护信息主体的合法权益；如果完全否定其合理性，又会对生成式人工智能的发展造成较大限制[9]。因此，第二十七条在生成式人工智能场景下留下了较大的解释空间，也增加了法律适用的不确定性。

实践中，已公开个人信息被用于人工智能训练而产生的侵权纠纷已不鲜见。在北京互联网法院审理的殷某某诉某智能科技公司等人格权侵权纠纷案中[2]，配音演员殷某某发现其声音被用于某 AI 配音软件的声音产品，播放量超过 32 亿次。法院认定，原告最初授权某文化传媒公司的范围仅限于录制有声作品，不包括对声音进行人工智能处理，被告在未获明确授权的情况下将原告声音用于 AI 产品开发构成侵权。这一裁判思路表明，“已公开”不等于“可任意使用”，公开数据的后续利用仍需考量原权利人的合理期待与人格利益保护，并非所有公开信息均可被无限制地用于 AI 训练和商业化开发。因此，如果将模型训练简单理解为已公开个人信息的合理利用，可能导致公开信息被无限度地利用，不利于保护信息主体的合法权益。

生成式人工智能模型通常依赖网络爬虫等技术大规模地收集互联网数据，其特点在于数据来源广、数量大，同时抓取的过程是完全自动化的。在实际抓取过程中，处理者通常是将网页上的全部内容作为一个整体来进行收集，而不是针对某一类信息进行精准筛选。因此，生成式人工智能自动获取的数据中往往同时包含着已经公开、尚未公开、敏感个人信息以及其他类型的数据。由于抓取行为具有批量化和自动化特点，信息处理者很难在数据收集阶段准确地识别不同信息的法律类型，也难以针对不同类型的信息分别适用相应的法律规则，例如，哪些属于可以依法利用的已公开个人信息，哪些属于依法应当受到更严格保护的敏感个人信息，在抓取阶段往往无法及时区分。这不仅增加了个人信息保护风险，也使第二十七条关于已公开个人信息处理的适用范围在实践中更加难以把握。

4. 个人信息保护制度的优化路径

(一) 构建分级同意机制

传统告知同意规则在生成式人工智能的问题上难以进行适用，在无法逐项履行告知义务、逐一取得信息主体同意的情况下，应当对现有的告知同意规则进行适当调整，在保障个人信息权益的同时兼顾人工智能技术发展的实际需求。具体而言，可以建立分级同意机制，根据个人信息的类型、敏感程度以及

[2]殷某某诉某智能科技公司等人格权侵权纠纷案，北京互联网法院(2023)京 0491 民初 12142 号民事判决书。

处理风险设置不同层级的同意要求，而非对所有个人信息一律适用相同的标准。对于敏感个人信息以及可能对人格权益造成重大影响的信息，仍应坚持单独同意、明示同意等较高标准，确保信息主体能够充分地知悉个人信息处理情况并作出决定；对于一般个人信息，可以结合具体的应用场景，允许采取一次授权、批量授权等方式，降低重复获取用户同意所带来的成本[10]；对于法律明确规定无需取得同意即可处理的个人信息，则依法适用相应的法定处理事由，无需再重复取得授权。

与此同时，应当改变传统的“一次告知、一次授权”的处理方式，将告知义务贯穿于个人信息处理的全过程。信息处理者应当及时向信息主体公开个人信息处理规则、模型训练用途、数据使用范围以及退出方式等重要内容，同时，为确保信息主体能够准确地掌握个人信息处理的相关情况，在当个人信息处理目的发生重大变化或者处理方式发生明显调整时，信息处理主体负有重新告知信息主体的义务，以便其依据法律法规行使撤销授权等权益。

(二) 引入“正当利益”条款

《个人信息保护法》第十三条规定的个人信息处理合法性基础，在生成式人工智能场景下存在一定局限，逐一取得信息主体同意不具有现实可行性，已公开个人信息规则又难以满足模型训练需求，且对于敏感个人信息的处理也缺乏明确的适用依据。

针对这一法律适用空缺，可以在《个人信息保护法》中引入借鉴域外的“正当利益”条款作为个人信息处理的合法性基础，并将其限定适用于符合一定条件的信息处理行为。对于以模型训练、算法优化等为目的处理个人信息的，在无法取得个人同意且不符合其他法定处理事由的情况下，可以允许信息处理者以正当利益作为处理依据，但不得当然适用，而应当设置严格的适用条件，防止个人信息被不当利用。对此，可以建立正当利益审查机制，对相关的信息处理活动进行必要性和利益平衡审查。信息处理者首先应当证明模型训练具有合法、正当的目的，并说明处理个人信息对于实现该目的具有必要性；其次，应当综合考虑处理信息的种类、处理方式、影响范围以及可能对个人权益造成的影响，对数据利用利益与个人信息权益进行平衡[6]。当个人信息处理可能对信息主体合法权益造成较大影响时，应当优先保护个人信息权益，不得以正当利益为由过分扩大个人信息处理范围。

与此同时，引入正当利益事由并不代表着个人信息保护标准的降低。在使用个人信息的过程中，信息处理主体仍需遵循法律法规，承担个人信息保护的相关责任，强化对训练数据的管控，实施数据的去标识化、匿名化等安全防护手段，并依照法律规定，确保信息主体能够行使查询、修正、删除等合法权益。对于敏感个人信息或者涉及未成年人个人信息的处理，应继续坚持更加严格的保护标准，原则上不得直接适用正当利益作为处理依据。

(三) 强化服务提供者合规义务

除对个人信息保护制度进行原则与理论上的调整外，对于服务的提供者而言，也应当强化其自身应当承担的合规义务。首先，应加强训练阶段的数据来源管理。服务提供者在收集训练数据时，应当建立数据来源审查制度，对数据获取方式、数据来源渠道以及是否包含敏感个人信息等情况进行记录，并保存相应的审查资料。对于来源不明或者存在合法性争议的数据，应当立即停止使用或者及时予以清理，确保训练数据来源合法、过程可追溯，为后续风险排查和责任认定提供依据。

其次，应完善数据处理阶段的技术保护措施。服务提供者应根据个人信息的类型和风险程度将其进行层级化分类，对于不同等级的个人信息，进行相应的保护管理。对于一般个人信息，可以采取去标识化等方式降低识别风险；对于敏感个人信息，则应加强加密存储、访问控制等安全措施，减少个人信息泄露和再次识别的可能性。同时，应建立数据安全风险监测机制，及时发现并处置个人信息处理过程中存在的安全隐患。

最后，应加强人工智能在实际应用阶段的输出管理，明确服务提供者对生成内容的纠错与标记义务。

生成式人工智能模型在生成内容时，可能会输出涉及个人信息的内容。对此，服务提供者应建立对于生成内容的监测机制和风险降低机制，对模型生成内容进行持续检测，及时发现并处理可能泄露个人信息的输出结果。对于可能发生或者已经发生个人信息泄露的，应当及时采取停止生成、删除相关内容、更新模型等措施，防止风险进一步扩大。当生成内容涉及个人信息且被证实为错误时，服务提供者应在合理期限内通过模型微调、关键词屏蔽、输出过滤等技术措施，抑制该错误信息的再次生成。此种义务可参照《中华人民共和国民法典》[3]第一千一百九十五条“通知–删除”规则的规范逻辑，但需根据生成式AI的技术特征作相应调整，生成式人工智能所产生的错误信息无法像网络侵权内容那样被简单地“删除”，而是需要通过技术手段阻断其再次输出。同时，服务提供者应依据《互联网信息服务深度合成管理规定》[4]第十六条的要求，对 AI 生成内容进行显著标识，特别是在涉及个人身份、职业、信用等敏感领域，应采用更为显著的提示方式，告知用户该信息由 AI 生成、可能存在不准确之处。标记义务的功能不在于杜绝错误，而在于为用户提供风险判断的基础信息。

5. 结语

如何在促进人工智能发展的同时有效保护个人信息权益，成为当前个人信息保护制度面临的重要课题。对此，本文围绕现行制度在适用过程中存在的主要问题，提出了完善个人信息保护制度理论体系以及强化服务提供者的义务履行等建议。生成式人工智能的技术更新仍在加速，法律制度也需要在实践检验中不断进行调整，在不断革新中实现个人信息权益保护与人工智能创新发展的协调统一。

参考文献

[1] 叶雄彪. 生成式人工智能背景下的个人信息保护: 范式转换与规则完善[J]. 法学家, 2025(4): 61-73+192.

[2] 武腾. 人工智能时代个人数据保护的困境与出路[J]. 现代法学, 2024, 46(4): 116-130.

[3] 付新华. 生成式人工智能对个人信息保护法的结构性冲击及应对[J]. 法学论坛, 2025, 40(6): 102-112.

[4] 李川. 生成式人工智能背景下个人信息的刑法规制困境与归责出路[J]. 华东政法大学学报, 2025, 28(6): 22-34.

[5] 王若冰. 生成式人工智能场景下已公开个人信息的弱保护[J]. 中国人民大学学报, 2025, 39(6): 75-87.

[6] 王苑. 人工智能预训练中大规模抓取个人信息的合法性困境与出路[J]. 中国法律评论, 2025(5): 66-78.

[7] 斜晓东. 风险与控制: 论生成式人工智能应用的个人信息保护[J]. 政法论丛, 2023(4): 59-68.

[8] 王歌雅, 尹懋锟. 生成式人工智能训练语料的个人信息保护[J]. 哈尔滨商业大学学报(社会科学版), 2026(3): 118-128.

[9] 张新宝. 生成式人工智能训练语料的个人信息保护研究[J]. 中国法学, 2024(6): 86-107.

[10] 朱晓峰, 袁子烨. 论生成式人工智能学习端的个人信息分级同意规则[J]. 西北工业大学学报(社会科学版), 2025(2): 136-143.

[3] https://www.court.gov.cn/zixun/xiangqing/233181.html

[4] https://www.moj.gov.cn/pub/sfbgw/flfggz/flfggzbmgz/202307/t20230705_482071.html