Advances in Social Sciences 社会科学前沿, 2026, 15(8), 270-280
Published Online August 2026 in Hans. https://www.hanspub.org/journal/ass
https://doi.org/10.12677/ass.2026.158660

Hans汉斯

生成式人工智能服务提供者侵权责任中的注意义务

王　妍

大连海事大学法学院，辽宁　大连

收稿日期：2026年6月27日；录用日期：2026年8月4日；发布日期：2026年8月12日

摘　要

生成式人工智能的颠覆性发展对以人类行为为中心的侵权责任框架构成了根本性挑战。其侵权形态呈现出“人机协同”、“直接生成”与“算法黑箱”等全新特征，导致传统过错责任、无过错责任及“通知－删除”避风港规则均面临适用困境。注意义务作为联结过错认定与责任承担的关键枢纽，成为破解这一困局的核心。文章认为，生成式人工智能服务提供者并非传统网络服务提供者，而是兼具“守门人”与“信息生产者”角色的新型主体。其注意义务的理论来源可追溯至传播学“守门人”理论与侵权法危险控制理论，其正当性则建立在权利义务相一致、法经济学预防效率及全球数字治理趋势之上。在制度构建上，应摒弃非此即彼的归责模式，转向构建一个贯穿事前防范、事中干预、事后处置的全流程、分层级注意义务体系。具体而言，服务提供者需承担基于现有技术水平的同业高度注意义务，其判定基准应综合考虑预见可能性、控制力、获益情况及行业特性。通过合理设定注意义务，旨在技术创新与权益保护之间寻求动态平衡，引导人工智能产业向善发展。

关键词

生成式人工智能，服务提供者，注意义务，侵权责任，避风港规则，守门人理论

Duty of Care in the Tort Liability of Generative AI Service Providers

Yan Wang

School of Law, Dalian Maritime University, Dalian Liaoning

Received: June 27, 2026; accepted: August 4, 2026; published: August 12, 2026

Abstract

The disruptive advancement of generative artificial intelligence poses a fundamental challenge to

文章引用: 王妍. 生成式人工智能服务提供者侵权责任中的注意义务[J]. 社会科学前沿, 2026, 15(8): 270-280.
DOI: 10.12677/ass.2026.158660

the tort liability framework, which has traditionally been structured around human conduct. The distinctive modalities of infringement emerging in this context—namely, human-machine collaboration, direct generation, and algorithmic black boxes—render the existing doctrines of fault-based liability, strict liability, and the notice-and-takedown safe harbor regime substantially inadequate. The duty of care, serving as the pivotal nexus between the determination of fault and the imposition of liability, lies at the heart of resolving this predicament. This article contends that generative AI service providers are not conventional internet service providers; rather, they assume a hybrid role as both "gatekeepers" and "information producers." The theoretical underpinnings of their duty of care can be traced to the gatekeeper theory in communication studies and the risk control theory in tort law, while its normative justification is grounded in the principle of congruence between rights and obligations, the preventive efficiency rationale of law and economics, and the imperatives of global digital governance. In terms of institutional design, a binary approach to liability allocation should be abandoned in favor of a comprehensive, multi-tiered duty-of-care system that operates across the entire continuum of pre-event prevention, in-event intervention, and post-event remediation. Concretely, service providers should be held to a high duty of care commensurate with the prevailing technological standards of the industry, the assessment of which should take into account foreseeability, degree of control, realized benefits, and sector-specific characteristics. By calibrating the duty of care appropriately, it is possible to strike a dynamic equilibrium between technological innovation and the protection of legal interests, thereby steering the development of the AI industry towards socially beneficial ends.

Keywords

Generative Artificial Intelligence, Service Providers, Duty of Care, Tort Liability, Safe Harbor Rules, Gatekeeper Theory

1. 引言

以 ChatGPT、Deepseek 等大模型的相继兴起为标志，生成式人工智能正迅速成为科学研究与社会生活等众多领域的关键工具。与传统人工智能只能进行分析、识别和预测不同，生成式人工智能的本质在于“创造”。但正因训练数据规模庞大且缺乏透明度、算法决策带有“黑箱”色彩[1]，以及生成过程本身具有动态性和难以完全预见的特点，生成式人工智能在便捷生活、提升生产效率的同时，不仅掀起了前所未有的侵权风险浪潮，涵盖知识产权侵权、人格权侵权、内容安全与公共秩序风险等，也引发了另一类间接风险，即用户对其输出的医疗、法律等专业内容寄予极高信任而带来的危害。生成式人工智能的侵权表现出与传统网络侵权截然不同的样态，这使得传统侵权规则遭遇了根本性的适用困境[2]。

面对这一全新的侵权形态，以人类行为为中心构建的传统侵权规则体系显得捉襟见肘。司法实践与学界探索从未停止，但无论是主张适用严格产品责任，还是回归一般过错责任，抑或是简单延伸针对传统网络服务提供者的“通知－删除”避风港规则，均因其无法契合生成式人工智能“人机协同”、“直接生成内容”及“算法黑箱”等技术特性而陷入结构性困境。

因此侵权法中的“注意义务”概念被推至幕前[3]。注意义务是判定行为人是否存在主观过错的客观化标准，它能通过弹性化的设定，在鼓励创新与保护权益之间寻求动态平衡。为生成式人工智能服务提供者设定何种程度、何种内涵的注意义务，已成为破解当前归责困境的核心命题。本文旨在系统论证：

生成式人工智能服务提供者为何应承担注意义务？其理论来源与正当性基础何在？其注意义务的来源有哪些？最终，应如何构建一套既符合技术理性又契合法理逻辑的、全流程、分层级的注意义务制度体系？通过对这些问题的深入探讨，期望为我国生成式人工智能的法治化治理提供有益的理论支撑与制度参考。

2. 生成式人工智能侵权规则的现实困境

生成式人工智能的崛起对传统侵权规则的适用造成深刻挑战。生成式人工智能的独特性使得以人类行为为中心、以清晰因果关系为预设的传统归责体系无法适用。本部分将首先剖析生成式人工智能侵权的独特性[4]，进而论证传统归责原则与规则(过错责任、无过错责任、避风港规则)的适用困境，在此基础上得出这些困境的实质为注意义务的制度空白的结论。

2.1. 传统侵权规则理论无法解决生成式人工智能侵权问题

由于训练数据来源的海量与不透明性，以及生成过程的不完全可控性，生成式人工智能在带来生活便利及提高生产的同时，不仅带来了诸多的直接法律风险，如知识产权侵权、人格权侵权、内容安全与公共秩序风险等，也引发了因用户高度信任其输出的医疗、法律等专业内容而造成的间接风险[5]。这些侵权呈现出与传统网络侵权截然不同的特征，给传统侵权规则带来了根本性的适用困境。

传统侵权规则理论中[6]，完整的侵权责任通常需要满足四个要件：归责原则、加害行为、损害后果、因果关系。其中，归责原则构成了侵权责任认定的逻辑起点和核心框架。归责原则是指确定民事主体承担民事责任的法律根据与标准，即在损害事实已经发生的情况下，为确定行为人对其行为所造成的损害是否需要承担民事赔偿责任的原则。它是整个民事责任制度的核心，贯穿于侵权责任法和合同法之中，对责任的构成要件、举证责任的分配、免责事由的确定以及损害赔偿的计算方法等均起着决定性作用，包括过错责任原则，过错推定原则，无过错责任原则与公平责任原则。然而，由于生成式人工智能侵权的特殊性，其无法直接适用传统侵权规则。

其一，生成式人工智能侵权的损害来源表现出明显的“人机协同”属性。生成式人工智能侵权内容的生成区别于传统互联网信息的呈现方式，不是服务提供者单方行为所致，也不是机器自发形成的产物，而是由用户输入指令，经其模型基于海量训练数据的概率推理后输出生成内容。在这一过程中，服务提供者的模型设计、数据选择、参数调整，用户的提示词输入、使用方式，以及生成式人工智能自身的“算法幻觉”与技术局限，共同构成了损害来源。这使得传统侵权法上“加害行为”的识别变得异常困难，多因一果或因果链条中断的情形成为常态[7]。

其二，生成式人工智能服务提供者直接参与内容的生成，使得信息呈现已突破传统“信息中介”的角色定位。在互联网形成发展早期和中期，网络服务提供者扮演的是信息存储中介或信息交互媒介的角色，其并非内容的直接生产者。然而，在生成式人工智能得以长足发展之后，生成式人工智能服务提供者扮演的是信息生产中心的角色。其对生成内容具有从数据来源到算法设计再到结果输出的全流程控制力，掌握着模型参数、数据来源与输出行为。这种角色转变意味着，其已不再是单纯的技术中立者，而是内容生产的积极参与者和风险源开启者。传统侵权规则基于“信息中介”定位所设置的较低程度的注意义务标准，已无法匹配其“信息生产者”的真实地位[8]。

其三，生成式人工智能呈现出显著的“算法黑箱”特征，导致过错与因果关系的证明面临技术壁垒。传统侵权法所规制的行为，其发生机制通常是可追溯、可解释的。然而，生成式人工智能依托深度学习模型运行，其内部决策过程对用户乃至监管者而言都是一个难以窥探的“黑箱”。对于受害人而言，要穿透算法黑箱，证明服务提供者在模型设计、数据选择等环节存在过错，证明生成内容与损害之间的因果关系，在技术上几乎不可能[9]。这种技术特性的差异，使得传统侵权规则所依赖的过错认定和因果关

系证明机制面临失灵的风险。

2.2. 生成式人工智能服务提供者侵权责任中的注意义务空白

正是由于上述结构性差异，学界对生成式人工智能服务提供者的归责原则产生了激烈争论，但这些争论本身也暴露了现有选项的不足，其核心在于缺乏一个能够弹性回应技术特殊性的注意义务框架。

有观点主张将生成式人工智能视为产品，对服务提供者适用无过错责任。其核心理由在于：服务提供者更了解产品，由其承担无过错责任可降低受害人信息成本，符合法经济学上的风险控制逻辑[10]。然而，这一路径面临多重障碍。其一，缺陷证明困难。由于算法黑箱，受害人几乎无法证明生成式人工智能存在何种设计或警示缺陷以及该缺陷如何导致了损害。其二，发展风险抗辩。服务提供者很容易依据《产品质量法》第 41 条[1]，以“当时科学技术水平尚不能发现缺陷”为由免责，这使得产品责任对受害人的保护大打折扣。其三，用户责任共担。在生成式人工智能侵权中，受害人的诱导性输入可能是损害发生的重要原因，产品责任模式无法恰当处理这种多方责任混合的场景。其四，过度抑制创新。使新兴技术产业承担过重的无过错责任，可能严重抑制行业发展，不符合国家战略。

另有观点主张适用一般过错责任[11]。其理由在于避免对服务提供者自由的不当限制，并敦促用户承担合理注意义务。但过错责任的核心困境在于举证责任的失衡。面对算法黑箱，受害人几乎无法举证证明服务提供者在模型设计、数据筛选、风险预警等环节存在主观过错。服务提供者的技术细节通常作为商业秘密保护，这使得受害人的权益保护在实践中沦为空谈，无法有效倒逼服务提供者采取风险防范措施。

还有观点试图将“通知－删除”避风港规则延伸适用于生成式人工智能[12]。但生成式人工智能的行为模式与此存在根本差异。生成式人工智能是内容的直接生成者，而非被动存储者。《生成式人工智能服务管理暂行办法》第 9 条[2]已将其明确定性为“信息内容生产者”。不仅如此，避风港规则在生成式人工智能场景下本身就会失灵：生成式人工智能通过“洗稿”式创作，使侵权内容难以被简单识别；同时，海量用户通知将导致服务提供者的人工审核成本急剧增加，制度运行的经济基础不复存在。因此，使生成式人工智能服务提供者简单适用避风港规则，将过分降低其注意义务标准，既不公平，也非效率[13]。

上述分析表明，生成式人工智能侵权规则面临的困境，并非某一具体规则的修补问题，而是整个传统归责体系的结构性失灵。从困境的实质来看，问题的核心不在于归责原则的抽象选择，而在于如何构建一套能够弹性反映生成式人工智能侵权特殊性的义务规则体系。在侵权法上，注意义务正是这一弹性调节机制的核心。注意义务是判定行为人是否存在主观过错进而构成侵权的关键因素。注意义务的程度越高，侵权责任就越接近无过错责任；程度越低，则越接近一般过错责任。通过注意义务的精细化、场景化设定，可以联通过错与无过错、事前预防与事后救济，弥合二元对立的归责体系。

因此，生成式人工智能侵权规则的现实困境，最终落脚于注意义务的制度空白。需要以生成式人工智能服务提供者的新角色、新技术能力为基点，构建一套贯穿事前、事中、事后全过程的注意义务体系。这正是本文后续部分将要重点探讨的核心命题。

3. 生成式人工智能服务提供者注意义务的理论来源与正当性基础

在民法中，注意义务[14]是指行为人在进行某种活动时，应当保持合理谨慎，预见自己的行为可能对他人权益造成伤害，并采取适当措施加以避免的法律义务。它是判断行为人是否存在过错的核心标准之一。生成式人工智能服务提供者的注意义务并非凭空创设，而是建立在深厚的理论基础和充分的正当性论证之上。理论来源回答了“这一义务从哪些既有学说和实践中衍生而来”的问题，正当性基础则回答

[1]https://flk.npc.gov.cn/detail?id=ff8080816f135f46016f1d6dfd7614d3&fileId=&type=&title=
[2]https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

了“为何必须赋予服务提供者这一义务”的问题。两者相互支撑，共同构成注意义务设定的完整理据。

3.1. 理论来源

生成式人工智能服务提供者注意义务的理论来源，可从侵权法原理和司法实践两个维度加以追溯。

侵权法内部的危险控制理论与利益平衡理论，为服务提供者注意义务的承担提供了直接的法理支撑。危险控制理论认为，开启或维持危险源者负有采取合理措施控制、防止危险的义务。生成式人工智能服务提供者既是风险的引入者(通过设计模型、收集数据)，也是风险的获益者(通过商业运营获利)，同时其相对于用户和权利人，对技术内核拥有绝对的控制力。因此，由其承担首要的注意义务，符合“风险与收益对等”、“控制者承担责任”的基本法理。利益平衡理论[15]则要求法律在多元利益冲突中寻求动态和谐。生成式人工智能服务涉及服务提供者的商业利益、用户的表达自由与创作自由、权利人的知识产权与人格权，以及社会公众的知情权与数据安全。过高的注意义务会抑制技术创新与产业发展，而过低的注意义务则会使权利保护沦为空谈。注意义务作为一种弹性工具，其设定过程本身就是对不同利益进行权衡与取舍的过程，旨在实现“促进创新”与“保护权益”之间的最佳平衡点。

司法实践中已出现对生成式人工智能服务提供者注意义务的初步探索，为理论构建提供了宝贵的经验素材。最具代表性的当属广州互联网法院审理的“奥特曼案”((2024)粤 0192 民初 113 号)[3]。在该案中，法院认定被告作为生成式人工智能服务提供者侵犯了原告的复制权和改编权，并明确指出，被告应承担“合理、可负担的特定注意义务”，包括建立投诉举报机制、提示潜在风险、标明显著标识等。法院并未简单适用传统的“通知－删除”规则，而是要求服务提供者承担一定的事前预防和事中控制义务，这标志着司法实践已开始突破传统避风港规则的局限，探索符合生成式人工智能特性的注意义务框架。这一判决虽然仍显初步，但为后续理论构建和规则设计提供了重要的实践坐标。

3.2. 正当性基础

如果说理论来源回应的是义务“从何而来”的知识谱系问题，正当性基础则回应义务“为何应当如此设定”的价值证成问题。生成式人工智能服务提供者注意义务的正当性，可从法理、现实和国际治理三个维度予以论证。

首先，从法理逻辑出发，权利义务相一致原则与禁止权利滥用原则为注意义务的设定提供了规范性根基[16]。生成式人工智能服务提供者在运营过程中享有广泛权利：从知识产权角度看，其拥有算法、模型和数据等相关权利；从商业自由角度看，其有权自主决定商业模式；从利益获取角度看，其通过用户付费、商业广告等途径获取巨大经济收益。基于“权利应负义务”的基本法理，服务提供者理应承担与上述权利相对应的义务，涵盖内容质量控制、法律合规、算法透明和用户权益保护等多个方面。与此同时，禁止权利滥用原则要求，若服务提供者未尽必要注意义务，将经营成本与侵权风险转嫁给权利人和公众，则构成对其权利的滥用。因此，对服务提供者设定注意义务，是矫正其权利与义务失衡状态、防止权利滥用的必然要求。

其次，从现实考量出发，传统避风港规则的失灵与维权困境的倒逼，构成注意义务设定的现实驱动力。在网络版权保护的传统框架下，“通知－删除”规则旨在平衡各方利益。然而，生成式人工智能的技术特性已从根本上动摇了这一规则运行的前提[17]。其一，侵权发现难：算法黑箱和“洗稿”式创作使侵权内容难以被识别。其二，证据固定难：生成式人工智能的输出具有随机性和不可重复性，用户退出界面后，侵权证据难以再现。其三，有效通知难：权利人难以定位侵权内容在数据库中的具体位置。其四，

[3]参见广州互联网法院(2024)粤 0192 民初 113 号民事判决书。

维权成本高：服务提供者与下游应用的联动进一步增加了监控难度。从法经济学视角看，当侵权预防的社会收益大于预防成本时，由控制成本最低的一方承担责任是有效率的。生成式人工智能服务提供者处于风险控制的枢纽位置，拥有最低的预防成本和最强的控制能力，由其承担注意义务是实现社会总成本最小化的理性制度选择。

最后，从宏观政策考量，顺应全球从严治理趋势，有利于促进产业健康有序发展。以欧盟《数字服务法案》和《数字市场法案》为典型代表，全球主要经济体正在将平台治理从事后应对提前到事前防范，为超大型平台设置一系列积极义务[18]。这一治理转向的深层动因在于，传统的事后救济手段已难以有效应对大规模、系统性的侵权风险。我国作为人工智能技术和产业的重要参与方，坚持发展与安全并重、促进创新与依法治理相结合的原则，设置明确、合理的注意义务标准，一方面能够为产业提供稳定的法律预期，引导企业“科技向善”，另一方面也能缓解未来大量著作权纠纷涌入法院导致的司法不堪重负问题。

3.3. 域外经验的比较与启示

生成式人工智能侵权治理是全球性难题，欧盟的立法探索与美国的司法实践为此提供了重要参照，其经验与局限共同构成本文注意义务方案的国际对话背景。

欧盟采取了公法监管主导的路径。《人工智能法案》于 2024 年正式生效[4]，其基于风险分级的框架将通用人工智能模型纳入规制，要求提供者履行透明度义务(如公开受版权保护训练数据的摘要)、版权合规义务及系统性风险评估义务。然而，该法案本质上是行政监管工具，其设定的合规义务并不直接等同于侵权法中的注意义务，且“合规即免责”的逻辑难以应对个案中过错认定的复杂性。与此同时，《数字服务法案》突破了传统避风港规则的被动回应模式，要求在线平台承担积极的尽职调查义务，并依平台规模与风险程度设置梯级化义务，其核心启示在于注意义务应从“通知－删除”的事后反应升级为“风险评估－主动预防”的系统治理。但 DSA 的规制对象仍是“信息中介”，而生成式人工智能服务提供者是内容的“直接生成者”，二者在风险控制能力上存在本质差异，故无法简单移植。

美国则主要依托司法诉讼探索责任边界。在汤森路透诉罗斯案(2025 年)中，法院认定未经授权使用受版权保护的训练数据构成侵权，确立了训练数据阶段须履行来源合法性审查的基础义务。在纽约时报诉 OpenAI 案中，法院不仅认定了直接侵权，更以被告“实质性帮助”最终用户为由支持间接侵权主张，并指出服务提供者在收到侵权通知后即具有“推定知情”，其注意义务层级随之提升。这些判例的积极意义在于明确了训练阶段的合规义务不可推卸、通知是注意义务升级的触发机制；但其局限同样明显——法院仍依附于传统帮助侵权框架，未能针对生成式人工智能“直接生成内容”的技术特性构建独立的责任标准，且个案裁判缺乏系统性。

上述域外经验为本文提供了三方面借鉴：风险分级的差异化思路、从被动回应转向主动治理的义务转型，以及通过证据披露与推定规则矫正举证责任分配。在此基础上，本文方案实现了三重超越。其一，将服务提供者明确定性为兼具“守门人”与“信息生产者”的新型主体，而非传统的“中介”或“合规义务人”，为注意义务的全面设定提供了更为准确的主体定位。其二，构建了覆盖“事前防范－事中干预－事后处置”的全流程义务体系，克服了欧盟立法偏重事前合规、美国诉讼囿于个案回应的碎片化缺陷。其三，通过“过错推定 + 比例责任”的归责机制与“技术可行性 + 经济合理性”的双重边界，为注意义务保留了适应技术迭代的弹性空间，避免“一刀切”标准对中小企业的过度压制。这一方案在借鉴域外制度智慧的同时，立足于我国产业实际与法律传统，力求在全球人工智能治理的法治对话中贡献具有

[4]https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

解释力和操作性的中国方案。

4. 生成式人工智能服务提供者承担注意义务的制度构建

前文的论述已经揭示：生成式人工智能服务提供者既非传统的信息中介，也非普通的产品生产者，而是一种兼具“守门人”与“信息生产者”角色的新型主体[19]。其注意义务的设定，不能简单移植既有的避风港规则，也不能抽象地诉诸一般过错责任或无过错责任。一个科学的制度构建，必须回答三个层层递进的问题：第一，注意义务的具体内容是什么？第二，注意义务的履行标准如何判定？第三，注意义务的边界在哪里？本部分将围绕这三个问题，构建一个逻辑自洽、层次分明的注意义务制度体系。

4.1. 注意义务事先、事中与事后的协同体系

注意义务的履行应当贯穿于生成式人工智能服务的全生命周期。按照风险发生的时间顺序，可以将其划分为事先防范、事中干预和事后处置三个相互衔接的阶段。这三个阶段并非孤立的义务清单，而是遵循“预防为主、干预为辅、处置为补”的风险治理逻辑。

4.1.1. 在事先防范层面以源头控制为核心的风险预防义务

事先防范是注意义务的第一道防线，其目的在于从源头上降低侵权发生的概率。与传统的“通知－删除”规则所体现的被动、事后反应不同，事先防范要求服务提供者主动介入技术开发与模型设计的早期阶段。具体而言，此项义务包含四个层次：(a) 数据合法性审查义务。生成式人工智能的“知识”完全来源于其训练数据。如果数据本身包含侵权内容、违法信息或未经授权的作品，那么模型输出的内容便极有可能“继承”这些瑕疵。因此，服务提供者必须对预训练数据的来源、内容和标注活动履行合法性审查。这包括：确保语料来源具有合法授权或可追溯性(来源安全)；过滤违法不良信息、侵犯他人知识产权的内容以及未获授权的个人信息(内容安全)；建立规范的标注规则，并对标注人员进行专业培训，防止因标注错误引入偏见或侵权风险(标注安全)。此项义务是注意义务体系的基石，因为事后补救无法弥补数据源头固有的缺陷。(b) 模型设计的合法性审查义务。算法模型并非价值中立的工具，其架构设计、参数设定与优化目标直接影响生成内容的合规性。服务提供者应当在模型设计阶段嵌入合法性约束：一方面，审查算法本身是否具有诱导或放任侵权行为的倾向；另一方面，确保训练数据的多样性，以避免模型产生系统性歧视或偏见。此外，服务提供者应当引入“反馈循环”机制，使模型能够根据后续的合规反馈进行自我修正。这项义务的本质是将法律规范“代码化”，使合规要求内嵌于技术系统之中。(c) 生成内容的预防控制义务。即便数据与模型均合规，生成式人工智能仍可能因“算法幻觉”或用户诱导而产出侵权内容。因此，服务提供者需要在内容生成的前、中、后各环节设置预防性控制措施。在生成前，设定关键词过滤与敏感词库；在生成中，采用自动化内容筛选系统对实时输出进行合规检测；在生成后，对高风险内容进行人工复核或标记。此项义务的意义在于，它承认了技术无法做到零风险，但要求服务提供者采取成本可控的合理措施来降低风险。(d) 用户行为的规范管理义务。用户是生成式人工智能服务的直接操作者，其输入指令的质量与合法性直接影响输出内容。服务提供者应当通过用户协议、使用指南和风险提示等方式，明确告知用户不得利用服务实施侵权行为。同时，对于存在持续输入恶意诱导指令、明显滥用服务的高危用户，服务提供者应当采取警告、限制功能乃至封禁账户等措施。此项义务体现了“共同责任”的理念：服务提供者无法控制用户的每一次输入，但有义务通过制度设计来引导和约束用户行为[20]。

4.1.2. 在事中干预层面以透明度为核心的披露与监控义务

事先防范无法完全杜绝侵权风险，因此在服务运行过程中，服务提供者还需履行一系列事中干预义务。这些义务的核心在于提升透明度和实时风险控制。(a) 算法决策基本逻辑的披露义务。“算法黑箱”

是权利人维权和监管机构审查的最大障碍。服务提供者应当在保护商业秘密的前提下，向用户和监管者提供关于模型运行原理的可理解性说明。这种说明不必涉及具体的源代码或参数细节，而应聚焦于模型的决策路径、生成依据以及影响输出的关键因素。披露的目的在于使外部主体能够判断服务提供者是否履行了基本的注意义务，从而缓解信息不对称[21]。(b) 生成内容来源与使用规则的披露义务。权利人之所以难以维权，很大程度上是因为其无法确定侵权内容的“来源”。服务提供者应当向用户说明生成内容所依据的训练数据概况(例如是否包含受版权保护的作品)，并明确告知生成内容的合法使用边界(例如是否可用于商业目的)。这项义务不仅有助于用户评估侵权风险，也为权利人初步锁定侵权源头提供了线索。(c) 模型局限性与潜在风险的披露义务。生成式人工智能并非全知全能，其在特定场景下可能出现事实错误、逻辑矛盾或专业性不足。服务提供者应当主动向用户披露这些局限性，并对输出内容的准确性进行合理提示。特别是在医疗、法律、金融等高风险专业领域，服务提供者必须附加显著的风险警示，明确告知用户生成内容不能替代专业人士的意见。这项义务的意义在于，它改变了用户对生成内容的“盲目信任”状态，促使用户以批判性态度审视输出结果，从而间接降低损害发生的可能性。(d) 高危用户行为的实时监控与干预义务。对于在事中阶段发现的高危用户(如一日内多次输入敏感、偏激、诱导性指令)，服务提供者应当拒绝回答相关问题，并对用户进行风险提示。如果用户行为情节严重，服务提供者应当暂停提供服务，并采取进一步处置措施。此项义务是事先预防义务的自然延伸，它要求服务提供者建立动态的、实时响应的风险监测机制[22]。

4.1.3. 在事后处置层面以救济为核心的响应与修正义务

即便服务提供者已尽到事先防范与事中干预义务，侵权内容仍有可能生成并传播。在此情况下，事后处置义务成为防止损害扩大的最后一道防线。(a) 建立健全投诉举报机制与及时反馈义务[23]。服务提供者必须设置便捷、公开的投诉举报渠道(如电话、邮件、在线表单)，并公布处理流程与时限。权利人一旦发现侵权内容，可以通过该渠道发出通知。服务提供者在收到有效通知后，应当及时受理、处理，并将处理结果反馈给投诉人。此项义务是“通知–删除”规则在生成式人工智能场景下的改良适用，但其功能已从传统的“免责条件”上升为一项独立的注意义务。(b) 对生成内容的及时处理义务。当服务提供者通过自查或权利人通知发现侵权内容时，应当立即采取删除、屏蔽、断开链接等必要措施，以防止损害的进一步扩大。对于重复侵权的用户，服务提供者应当采取账号封禁等更为严厉的措施[24]。此项义务的履行标准是“及时性”与“有效性”，即措施应当在合理时间内采取，并能够实际阻断侵权内容的传播。(c) 合规性审查与模型的动态修正义务。事后处置不能止于个案救济，服务提供者还应当从每一次侵权事件中汲取经验，反哺模型的优化与规则的完善。具体而言，服务提供者应当定期对平台上的生成内容进行合规性审查，并根据审查结果、用户反馈以及法律政策的变化，动态修正模型参数、过滤词库和审核规则。这项义务使注意义务从“静态合规”走向“动态治理”，形成“风险发生–处置–学习–修正”的闭环。

4.2. 注意义务的判定标准

有了具体的义务内容，还需要一套科学的判定标准来回答“服务提供者是否尽到了注意义务”这一核心问题。判定标准应当兼顾客观层面的风险控制可能性与主观层面的理性人基准。

4.2.1. 客观判定要素

损害后果的可预见性。注意义务的成立以行为人对损害后果具有“可预见性”为前提。如果服务提供者客观上无法预见其模型可能产生某类侵权内容，那么法律便不能强求其采取预防措施。可预见性的高低，与被侵害权益的类型密切相关：对于著作权等法律明确保护且易于识别的权益，可预见性较高；

对于抽象的人格利益或不确定的新兴权益，可预见性较低[25]。因此，判断服务提供者的注意义务程度，首先需要考察其在当时的技术条件和知识水平下，是否能够合理预见到侵权行为的发生。

服务提供者与被控侵权行为的近邻性。所谓“近邻性”，是指服务提供者与被控侵权行为之间是否存在紧密的关联，从而使其具备控制风险的能力。如果服务提供者在收到权利人通知后，已经对被控侵权用户产生了“特定的注意义务”，那么后续发生的重复侵权行为便会因这种近邻性而提高服务提供者的注意义务标准。换言之，“通知”不仅启动了个案的处理程序，也提升了服务提供者对同一用户或同类行为的注意层级。

公平正义的政策考量。注意义务的设定并非纯粹的技术判断，还蕴含着价值权衡。如果要求服务提供者承担过高的注意义务，将导致其运营成本剧增，抑制技术创新；反之，如果注意义务过低，则会使权利人维权无门。因此，在具体案件中，法院需要综合考量产业发展阶段、技术成熟度、社会公共利益等因素，动态地确定注意义务的合理边界。

4.2.2. 主观判定基准

从主观层面看，注意义务的违反以行为人未能达到“理性人”的注意程度为标志。在生成式人工智能领域，这一“理性人”标准需要进行行业化重构。

“理性人”是具有普遍谨慎与预见能力的同行业服务提供者。此处不应以普通人的注意水平为参照，而应以一个具备生成式人工智能领域基本技术知识、遵循行业惯常做法的“中等水平”服务提供者为基准。这意味着，初创企业和非专业技术提供者不能以“我不知情”为由降低注意义务，而应当参照行业通常的技术和管理水平来履行义务。

理性人标准应体现行业专业性。生成式人工智能涉及数据挖掘、模型训练、算法更新等一系列高度专业化的活动。因此，判断服务提供者是否达到理性人标准，需要考察其是否利用其所掌握的“专业技术优势”来识别和预防风险。例如，一个拥有先进算法团队的头部企业，其理性人标准应当高于一个仅提供简单 API 接口的小型服务商。

理性人标准具有动态演进性。随着技术的迭代和行业共识的形成，原本属于“高水平”的注意义务可能逐渐演变为“平均水准”。例如，在生成式人工智能发展初期，建立投诉举报机制可能被视为一项“额外”的注意义务；而在“奥特曼案”之后，这一机制已经成为行业标配。因此，理性人标准应当与时俱进，不能固守某一时点的技术水平。

4.3. 注意义务的合理边界

在确定了注意义务的内容与判定标准之后，还必须为其划定一个清晰的边界，以避免“过犹不及”。注意义务的边界主要由两个维度构成：技术可行性维度与经济合理性维度。

4.3.1. 现有技术水平作为技术可行性边界

“法律不强人所难”。如果一项风险预防措施超出了现有技术的可实现范围，那么即使该措施在理论上能够降低侵权风险，服务提供者也不因未采取该措施而承担法律责任[26]。例如，在现有技术下，完全消除“算法幻觉”是不可能的，因此不能要求服务提供者保证模型永远不会输出错误信息。同样，对于尚未被任何现有过滤技术识别的全新侵权模式，服务提供者未能预见和阻止该侵权，也不应轻易认定其违反注意义务。

4.3.2. 同业平均注意水平作为经济合理性边界

即使某项措施在技术上可行，如果其成本远远超过可能避免的损害，那么要求所有服务提供者一律采取该措施可能是不经济的。因此，注意义务的标准应当以该行业内“通常的、平均的”注意水平为参

照，而非以最优者的水平为标杆。这一参照标准为不同规模、不同技术能力的服务提供者提供了相对公平的评判尺度，避免了“一刀切”的高标准对中小企业造成过度压制。

4.3.3. 场景化与类型化的差异化边界

注意义务的边界还应因服务模式、风险等级和用户类型的不同而有所差异。原则上，服务提供者对内容的控制力越强、从服务中获益越多、服务所涉领域的风险越高，其注意义务的边界就越应向外延伸。反之，对于一个仅提供开源模型接口、不直接面向终端用户的底层模型开发者，其注意义务的边界应限于模型本身的安全性审查，而不应过度延伸至用户的具体使用行为。

4.4. 违反注意义务的法律后果

最后，还需要明确违反注意义务的法律后果。在生成式人工智能侵权的诉讼中，应当采用“过错推定与举证责任倒置”的证明机制。

4.4.1. 初步证据的提供与过错的初步推定

权利人只需证明：(1) 其享有受法律保护的权益；(2) 生成式人工智能生成的内容侵犯了该权益；(3) 该侵权内容是由被告提供的服务所生成。完成这三项初步证明后，法律应推定服务提供者存在过错(即未尽到合理注意义务)。

4.4.2. 服务提供者的反证机会

服务提供者可以通过举证推翻上述推定，证明其已经履行了与现有技术水平和同业平均标准相适应的注意义务。具体而言，服务提供者可以提供：(1) 数据合法性审查记录；(2) 模型设计合规性证明；(3) 内容预防控制措施的运行日志；(4) 用户行为管理记录；(5) 投诉举报机制的运行情况；(6) 对侵权内容的及时处置记录等。如果服务提供者能够证明其已尽到前述(一)至(三)部分所要求的各项义务，则应认定其不存在过错，不承担侵权责任。

4.4.3. 注意义务履行程度与责任比例的关联

即便服务提供者未能完全证明其无过错，法院也可以根据其履行注意义务的程度，相应地减轻其责任比例。例如，如果服务提供者已建立投诉机制，但处理不够及时，法院可以据此判定其承担次要责任而非全部责任。这种比例责任的引入，能够更精细地反映过错程度与损害后果之间的因果关系。

参考文献

[1] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. and Pedreschi, D. (2018) A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, **51**, 1-42. https://doi.org/10.1145/3236009

[2] 曹然. 从“避风港”到“认养池”——生成式 AI 智能体侵权的平台责任转型[J]. 南京社会科学, 2026(4): 121-132.

[3] 高阳. 通用人工智能提供者内容审查注意义务的证成[J]. 东方法学, 2024(1): 189-200.

[4] 张华韬. 动态体系论下生成式人工智能侵权的归责与构成[J]. 法学杂志, 2025, 46(6): 82-101.

[5] 黄锫. 人工智能大模型训练数据的风险类型与法律规制[J]. 政法论丛, 2025(1): 23-37.

[6] 杨立新. 侵权责任法[M]. 第 5 版. 北京: 法律出版社, 2025.

[7] 林洹民. 论人工智能致损的特殊侵权责任规则[J]. 中外法学, 2025, 37(2): 344-362.

[8] 冯晓青, 郭畅. 生成式人工智能平台著作权侵权责任与分层治理——基于杭州奥特曼案的思考[J]. 数字法治, 2025(2): 56-75.

[9] 张新宝, 卞龙. 论生成式人工智能服务提供者的过错推定责任[J]. 北方法学, 2025, 19(5): 5-19.

[10] 徐伟. 生成式人工智能服务提供者侵权归责原则之辨[J]. 法制与社会发展, 2024, 30(3): 190-204.

[11] 丁晓东. 从网络、个人信息到人工智能: 数字时代的侵权法转型[J]. 法学家, 2025(1): 40-54+191-192.

[12] 刘维. 论数字时代“避风港规则”的守正与创新[J]. 南大法学, 2026(1): 139-153.
[13] 司晓. 网络服务提供者知识产权注意义务的设定[J]. 法律科学(西北政法大学学报), 2018, 36(1): 78-88.
[14] 王利明. 侵权责任法研究(上、下卷) [M]. 最新版. 中国人民大学出版社, 2016.
[15] 冯晓青. 知识产权法利益平衡理论[M]. 北京: 中国政法大学出版社, 2006.
[16] 彭诚信, 苏昊. 论权利冲突的规范本质及化解路径[J]. 法制与社会发展, 2019, 25(2): 72-88.
[17] 朱晓峰. 生成式人工智能个人信息侵权因果关系不明时的责任认定[J]. 政法论坛, 2025, 43(6): 18-33.
[18] 吴沈括, 胡然. 平台治理的欧洲路径: 欧盟《数字服务法案》《数字市场法案》两项提案分析[J]. 中国信息安全, 2021(1): 71-74.
[19] 吴汉东, 樊赛尔. 试论生成式人工智能服务提供者的合理注意义务[J]. 知识产权, 2025(11): 3-23.
[20] 张建华. 信息网络传播权保护条例释义[M]. 北京: 中国法制出版社, 2006.
[21] 崔国斌. 网络版权内容过滤措施的言论保护审查[J]. 中外法学, 2021, 33(2): 305-326.
[22] 杜娟. 德国《版权服务提供商法案》的解读及对我国的借鉴[J]. 科技与出版, 2020(11): 91-98.
[23] 解正山. 约束数字守门人: 超大型数字平台加重义务研究[J]. 比较法研究, 2023(4): 166-184.
[24] 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.
[25] 苏宇. 大型语言模型的法律风险与治理路径[J]. 法律科学(西北政法大学学报), 2024, 42(1): 76-88.
[26] 严驰. 生成式 AI 大模型的全球治理方案: 何以可能与何以可为[J]. 科技与法律(中英文), 2024(1): 91-99.