

生成式人工智能服务提供者的注意义务探析

刘光晓

山东师范大学法学院, 山东 济南

收稿日期: 2026年6月3日; 录用日期: 2026年6月22日; 发布日期: 2026年7月13日

摘 要

认定生成式人工智能服务提供者是否侵权, 关键在于其是否尽到合理的注意义务。现行法律没有明确规定服务提供者的注意义务, 学界也对此看法不一。本文采用过错责任原则, 把注意义务分成了三个阶段: 风险预防、风险控制、风险处置。每个阶段各有三项义务。预防阶段是数据合法性审查、模型设计合规和用户管理。控制阶段是内容标识、风险提示和算法透明。处置阶段是投诉举报机制、内容阻断和动态修正。AI服务用户数量庞大、应用场景多样。AI的服务模式、应用场景、主体类型和用户群体等因素对注意义务都有影响, 因素变化了, 注意义务也会有所差异。该分类框架, 有助于更客观地判断服务提供者是否有过错。

关键词

生成式人工智能, 服务提供者, 注意义务, 类型化, 过错责任

Analysis of the Duty of Care for Providers of Generative Artificial Intelligence Services

Guangxiao Liu

Law School, Shandong Normal University, Jinan Shandong

Received: June 3, 2026; accepted: June 22, 2026; published: July 13, 2026

Abstract

Determining whether a generative AI service provider has infringed depends crucially on whether it has fulfilled its duty of reasonable care. Current laws do not clearly define the duty of care for service providers, and academic opinions remain divided on this issue. This article adopts the principle of fault liability and divides the duty of care into three stages: risk prevention, risk control, and risk response. Each stage entails three specific obligations. The prevention stage includes reviewing data legality, ensuring compliance in model design, and managing users. The control stage

involves content labeling, risk warnings, and algorithm transparency. The response stage covers mechanisms for complaints and reports, content blocking, and dynamic adjustments. Given the vast number of AI service users and diverse application scenarios, factors such as service models, application contexts, types of entities involved, and user demographics all influence the duty of care—changes in these factors lead to variations in the required level of care. This classification framework helps achieve a more objective assessment of whether a service provider has acted negligently.

Keywords

Generative Artificial Intelligence, Service Provider, Duty of Care, Typification, Fault Liability

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

生成式人工智能是互联网飞速发展的产物，是人工智能的重要分支，它可以根据算法、模型等生成相应的文本、图片、声音等。2025年7月，杭州互联网法院审结了一起案件。该案件是一起因AI“幻觉”引发的侵权纠纷案件¹。该案中，原告使用某公司的AI服务查高考志愿，AI给出了错误信息，并声称若信息错误将赔偿十万元。原告相信了，用该信息填报了志愿。后来原告发现AI提供的信息是错的，于是起诉至法院，要求该AI公司赔偿9999元。法院驳回了原告的诉讼请求。理由有二，一是AI生成的信息不是法律上的意思表示，不产生效力。二是AI侵权应适用过错责任原则。若服务提供者尽了合理的注意义务，没有过错则无需赔偿。这一案件引发了公众讨论，引出了一个问题：生成式人工智能服务提供者的注意义务，到底该如何认定？

2023年出台了《生成式人工智能服务管理暂行办法》(后文简称《暂行办法》)。这个办法要求服务提供者对AI生成的内容负有管理义务，但这些规定主要停留在行政监管层面。在民事侵权中，注意义务的具体内容并不明确。在注意义务的内容上，有学者从风险的预防、控制与处置三个环节构建注意义务框架[1]，也有学者从技术特性、非技术和程序性三个维度划分注意义务[2]。

关于归责原则，本文主张采用过错责任原则。与严格责任原则和无过错责任原则相比，过错责任原则更加符合生成式人工智能的技术特性，该技术具有算法黑箱、输出内容随机性、幻觉难以消除等特性。这些特性使得服务提供者无法精准控制每次生成的内容。严格责任原则要求服务提供者对无法预见、难以控制的侵权后果承担绝对责任，可能在一定程度上会抑制技术创新；无过错责任原则虽然可以减少受害人的举证负担，但是在AI幻觉随机性强、因果关系复杂的场景下，容易导致责任泛化。而过错责任原则要求服务提供者在其技术能力和行业通常水平范围内履行合理义务，不要求其承担所有风险责任。其次，从信息经济学视角来看，服务提供者处于信息优势地位，其承担证明已尽到注意义务的责任，可以降低信息不对称导致的交易成本，鼓励其主动采取防范措施。风险社会理论也支持这一立场，其认为法律不应该追求消除所有技术风险，应该通过合理分配注意义务，实现风险控制和产业发展的平衡。基于此，本文以过错责任原则为基本立场，对服务提供者的注意义务进行类型化分析。

本文将从风险预防、风险控制、风险处置三个阶段来分析注意义务。第一部分主要探讨注意义务的

¹梁某诉某人工智能公司侵权案，杭州互联网法院(2025)浙0192民初18143号民事判决书。

产生基础，第二部分阐述选择这一框架的理由，第三部分介绍注意义务的具体内容，第四部分讨论不同情况下的义务差异。

2. 人工智能服务提供者承担注意义务的原因

生成式人工智能服务提供者承担注意义务，主要有三个理由。第一，服务提供者是人工智能产生风险的“开关”。生成式人工智能存在固有的“幻觉”问题，即使经过服务提供者合规训练，但由于用户可能输入的提示文本过于复杂和庞大，甚至恶意输入误导性文本，其仍会生成错误或虚假性信息。服务提供者是这一技术的主要负责人，将这一技术推向市场，供不特定用户使用，实质上是开启了一个可能造成损害的风险源头。根据侵权责任编中的基本原理，开启或维持危险源的人，应当采取合理措施控制和防范风险，因此，服务提供者应承担相应的注意义务。

第二，在生成式人工智能技术的产生过程中，服务提供者既是提供信息交互服务的主体，内容生成服务的提供主体，发挥关键作用，是连接模型开发者、数据提供者和终端用户的关键一环。与其他主体相比，服务提供者对模型的运行、处理数据的方式以及生成内容的逻辑更加了解，具有更强的决定权和支配权。有学者将服务提供者这一角色称为“守门人”[3]，即谁开启控制危险源，谁就负有防范处置危险的义务和责任。正因服务提供者的特殊地位，其对 AI 模型生成的内容预见性更强，最有可能避免发生危险，所以其应承担相应的注意义务。

第三，服务提供者的注意义务，是在一般过错责任指导下合理的注意义务，是有限度的。不是承担无限责任或无过错责任。原因在于注意义务的设定应在提供服务的成本和收益之间达到平衡，且应在客观条件即现有技术基础之上设定义务。如果预防风险发生的措施成本远高于可能造成的损害，则不应苛求服务提供者必须采取此预防措施。服务提供者应采取本行业中通行的防范措施，在成本和收益之间寻求平衡，这既保护了用户的合法权益，也为服务提供者的技术创新保留了合理空间。

3. 注意义务类型化框架的选择

现行法律法规对生成式人工智能服务提供者的注意义务虽有原则性规定，但并未明确列出其内容。《中华人民共和国民法典》(后文简称《民法典》)《暂行办法》《中华人民共和国网络安全法》等有所涉及，为司法实践提供了参考，但法官的自由裁量空间仍然较大。

目前学界对提供者注意义务通过一定的标准进行了划分，主要可概括为两种研究路径。第一种是以时间为标准，将注意义务划分为事前的风险预防、事中的风险控制和事后风险处置；另一种是从功能属性上划分，分为技术特性、非技术和程序类三类注意义务。这两种研究路径各有其优点，本文选择前者，依据前者的分类标准进一步探析注意义务的具体内容。

本文采用事前风险预防、事中风险控制和事后风险处置的分类标准，是因为这与生成式人工智能技术的运行阶段相适应。服务提供者在将人工智能推向用户之前，应对人工智能所用到的数据、可能收到的语料以及其运行模型进行事前审查等，提高危险的可预见性，有效预防风险的发生；在 AI 技术运行过程中，如发现用户恶意输入不良内容诱导 AI 生成相应内容时，应及时干预，这要求提供者配备及时准确的监测系统等措施，能够及时控制即将发生的风险；风险发生后，提供者需要完善 AI 技术，不可视而不见。具体包括完善相应用户举报、投诉机制等措施，及时修改处置。此外，还有学者从 AI 技术的生命周期角度出发，将 AI 模型分为了模型训练、运行应用、优化迭代三阶段，主张递进式[4]。这一划分与本文的研究框架高度一致，模型训练对应风险预防，运行应用对应风险控制，优化迭代对应风险处置，这也为本文的框架选择提供了另一层理论支撑。有学者指出，生成式人工智能服务提供者是一种新型信息服务主体，与传统网络服务提供者有着本质区别[5]。这种对应关系能够更加清晰直观地展示这种特殊提供

者的注意义务，便于在司法实践活动中根据人工智能的运行阶段审查服务提供者是否尽到了相应的注意义务。

《生成式人工智能服务管理暂行办法》²分别从数据安全、内容安全、投诉处置等环节规定了服务提供者的义务，这些章节结构与“预防－控制－处置”的框架相一致。因此，本文以该框架作为分类标准展开分析。需要说明的是，注意义务的具体内容以及强度不是固定不变的，有学者提出应根据防范风险的成本、风险的识别难度以及现有的技术水平等动态因素加以判断^[6]，这一思路在下文中具体义务部分也会有所体现。

4. 注意义务的具体内容

与传统网络产品相比，生成式人工智能服务具有显著特殊性。其一是传统网络产品主要是提供信息检索、搜集和储存服务，为用户提供的内容主要是从数据库中直接抓取的，并非根据用户需求主动创造生成的；而生成式人工智能服务主要功能是根据用户输入的词条生成全新的内容，其生成结果不是预先存在数据库中，而是通过一定算法实时计算产生的；其二是传统网络产品的运行模式相对透明，预见性强，平台可根据用户行为制定相对明确的规则进行约束，能够有效预防风险的发生；而生成式人工智能自身存在“幻觉”和“黑箱”缺陷，其生成的内容会根据不同语料而变化，容易生成错误或者误导用户的信息，难以控制；第三，《民法典》第1195条中对传统网络产品的侵权责任认定适用“通知－删除”规则，平台在收到通知后进行删除、屏蔽等措施即可免责；而生成式人工智能发生侵权后，不能仅通过删除的措施免责，仅采取删除措施不足以免责，还需要采取阻断、过滤等技术措施。这些特殊性决定了生成式人工智能服务提供者的注意义务不能简单使用传统网络产品的规则，需要根据其特殊性，在风险预防、风险控制和风险处置框架下进行具体分析。

服务提供者采取事前预防，是人工智能服务平稳运行的前提。在风险预防层面，主要义务如下。第一种是数据合法性审查义务，包括语料来源是否合法、数据中是否含有违法信息以及是否侵犯了他人知识产权等，增强数据的真实性、准确性。第二种是模型设计合法合规义务，设计符合安全伦理要求，防止算法产生歧视、虚假或者违法内容。《暂行办法》第4条中有明确规定，具体要求服务提供者建立算法安全审查机制、设置关键词过滤机制等。第三种是用户行为管理义务。《暂行办法》第14条规定了对反复输入诱导性内容的用户，服务提供者应将其标记为高危用户并采取相应措施，避免侵害他人合法权益。这些事前预防措施旨在从源头上减少侵权风险，是服务提供者注意义务的第一道防线。有学者从“守门人”理论出发，论证了服务提供和承担事前预防义务的正当性。

事前的风险预防虽然重要，但是无法覆盖所有风险。当人工智能开始运行时，服务提供者还需要履行风险控制的注意义务。第一个是内容标识义务。《暂行办法》第12条、《人工智能生成合成内容标识办法》³中都明确了该义务，要求服务提供者对AI生成的内容做显著标识。第二个是风险提示义务，要求服务提供者在用户协议中以显著的方式向用户提示，AI生成内容可能会存在错误等，提醒用户谨慎采用，甄别生成内容。特别在医疗、法律、教育等专业领域，需要更显著的方式强化提醒，明确告知用户AI提供的内容不可代替专业意见。有学者进一步将风险提示义务细化成三方面，在使用前，通过用户协议和界面告知用户生成内容可能存在风险或错误，使用过程中，设置动态提示提醒用户生成内容可能和预期不一致，生成后，提示用户生成内容可能涉及第三方版权，建议在商业使用前进行合规审查。第三个是算法透明义务，要求服务提供者在不泄露商业秘密的同时，向用户说明该AI算法的基本原理和运行

²https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm

³https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm

逻辑，使用户更加了解，保障其知情权^[7]。总之，风险控制层面的注意义务，要求服务提供者对用户进行有效提示和必要披露。

若服务提供者尽到了预防和控制义务，侵权风险还是可能发生。这时候，需要事后处置层面的注意义务来补充。第一，建立投诉举报机制，要主动向用户公布处理流程和反馈时限。《暂行办法》第15条明确规定了这项义务。2024年，广州互联网法院宣判了“奥特曼”案。⁴某AI网站因生成了与“奥特曼”形象相似的图片，被判侵权。法院认定AI公司侵犯了原告的复制权和改编权。公司没有建立投诉举报机制，未尽到合理的注意义务。第二，内容阻断义务。服务提供者发现AI生成侵权内容后，要及时删除。此外，还需要采取技术手段，如关键词过滤、模型微调等，防止同样的内容再次生成。在前述“奥特曼”案判决中，法院也明确指出服务提供者有阻断义务。第三，动态修正义务。注意义务不是一成不变的。它要根据用户反馈和相关法律法规的变动进行调整。定期对AI模型测评和合规审查，查缺补漏。有学者指出，AI服务在优化迭代阶段，服务提供者有“通知-处置”的义务。跟传统“通知-删除”义务不同。“通知-处置”不仅要删除侵权内容，还要对模型诊断优化，从根本上杜绝同类侵权风险。总之，在这一层面的注意义务，就是要求服务提供者在侵权发生后能够及时阻断、持续改进。

上文提及的内容审核和用户管理等义务，虽然能够防范侵权风险，但是也存在一定局限性，主要表现为限制用户的言论自由、信息获取权以及阻碍技术创新等。内容审核机制过于严苛，可能会使得一些合法、有益的生成内容被屏蔽，抑制用户的正常表达需求；监控风险时，过度收集用户的行为数据可能会侵犯用户的个人隐私。因此，服务提供者在履行注意义务时，不应片面追求安全最大化，而应该在确保安全与保护基本权利之间寻求合理平衡。具体而言，可以遵循以下原则：一是比例原则，审核内容的强度应该与侵权风险的程度相匹配，对明显违法、不合理的内容进行严格管控，对一般性争议内容进行轻度干预或者事后处置；二是透明度原则，服务提供者应公开内容审核的基本规则和管理用户行为的标准，让用户了解其权利边界；三是救济机制，为用户建立健全申诉渠道，可以在有效防范侵权风险的同时，最大程度地保障用户的言论自由和信息获取权，也在一定程度上促进技术创新。

5. 注意义务的差异化适用

服务提供者注意义务的分类，不只是对义务内容分类。同一项注意义务，在不同场景、对不同人群，其强弱也不同。AI服务用户众多，使用场景很广泛，技术更迭快。在这种背景之下，AI服务会遇到很多新状况。所以，注意义务不能一刀切。应按照服务模式、应用场景、主体类型等因素调整，灵活适用。

5.1. 有偿服务与无偿服务的区分

常见的AI服务有两种，有偿服务和无偿服务。在有偿服务中，用户付了钱，对生成内容有合理期待。服务提供者因此获取利益，就需要承担更高的注意义务。例如：强化信息核验、建立责任追溯机制、为用户提供更便捷的反馈渠道等。无偿服务中，用户是免费试用的。用户对生成内容的准确性，不宜过高。因此，服务提供者的注意义务可以适当降低一些。杭州“奥特曼”案中，AI服务是有偿的。⁵法院判决被告因从AI创作中获取了直接利益，所以应承担较高的注意义务。

5.2. 专业场景与通用场景的区别

AI技术的快速发展，使其广泛渗透到社会的多个领域，例如自动驾驶、AI医生、智能客服、AI

⁴某上海文化发展有限公司诉某广州网络科技有限公司网络信息侵权案，广州互联网法院(2024)粤0192民初113号民事判决书。

⁵某上海文化发展有限公司诉某杭州智能科技有限公司著作权侵权及不正当竞争纠纷案，杭州互联网法院(2024)浙0192民初1587号民事判决书。

风险评估、智能物流、智能教育等。我们可以将生成内容的专业化程度作为标准，将 AI 的应用场景分为专业场景和通用场景。专业场景主要包括医疗、法律、教育等，其生成内容专业化程度高，一旦出错就可能对用户造成极大损害，因此服务提供者应承担更高的注意义务。具体包括强化“非专业建议”的标识、使用更权威的数据库以及采用多模型交叉验证等，降低风险。通用场景主要是指日常信息的查询和生成娱乐内容等，重点在于基础合规和适当提示，不要求服务提供者承担超出行业通用标准的义务。

5.3. 不同主体类型的区别

AI 服务的研发和运行，需要多个主体共同参与。基础模型开发者发挥关键作用。他和服务提供者的注意义务不同。前者主要是对模型的核心安全负有更高的注意义务。在数据训练、设计算法阶段做合规措施，例如对数据的合法性审查、算法安全评估等。后者主要负责管控即时内容和提示用户，如内容标识、风险提示和投诉举报处理等。

5.4. 对未成年人等特殊用户的保护

未成年人身心还没发育成熟，需要重点保护。《中华人民共和国未成年人保护法》规定，在网络环境中应保护未成年人的隐私和个人信息，防止其受到不良信息侵害。在 AI 服务中也一样，要给予未成年人更多保护。《暂行办法》第 10 条要求服务提供者采取有效措施，防范未成年人过度依赖或沉迷 AI 服务。具体内容为：第一，AI 产品设计和服 务要考虑未成年人的特点；第二，建立用户身份识别机制，如发现是未成年人，立即采取特殊保护措施；第三，限制使用时长，过滤不合适的内容；第四，协助家长监督管理孩子的使用行为。

综上，注意义务的差异化配置，就是在坚持基本义务框架的前提下，服务模式、应用场景、主体类型和用户群体等因素灵活调整。在保护权益和促进产业发展之间寻求平衡。

6. 结语

本文在过错责任原则指导下，以“风险预防、风险控制和风险处置”为框架，对注意义务进行了类型化分析。本文的基本观点可以概括为三点。第一，注意义务的产生是因为服务提供者开启了风险、有能力控制风险，且需要在控制成本和收益之间找平衡。第二，注意义务的内容可以分为三层：风险预防层面包括数据审查、模型设计、用户管理；风险控制层面包括内容标识、风险提示、算法透明；风险处置层面包括投诉举报机制、内容阻断、动态修正。第三，注意义务并非一成不变，而是根据服务类型、使用场景、用户群体等不同而变化。

本文的主要贡献在于两点。一是把现有研究的分类框架做了分析整合，按 AI 技术运行阶段重新梳理，便于认定注意义务；二是对注意义务的差异化配置进行比较系统的讨论，弥补了现有研究在这方面的不足。从司法实践来看，法院已经在个案中使用注意义务来判断服务提供者是否有过错。杭州“幻觉”案和广州“奥特曼”案都说明了，注意义务的认定要结合具体案例综合判断。

本文也有一定局限性，主要问题是案例不多，实证检验不充分。以后可以用更多案例来检验这个框架，也可以做比较法研究，借鉴国外相关的成功经验。

参考文献

- [1] 毕文轩. 论生成式人工智能服务提供者的注意义务[J]. 苏州大学学报(哲学社会科学版), 2025, 46(2): 90-103.
- [2] 姚佳. 生成式人工智能服务提供者的注意义务[J]. 法学论坛, 2026, 41(3): 17-31.
- [3] 杨利华. 生成式人工智能服务提供者注意义务研究[J]. 比较法研究, 2025(3): 54-68.

-
- [4] 李丹. 生成式人工智能服务提供者的阶段性注意义务与侵权责任[J]. 现代法学, 2026, 48(2): 224-242.
 - [5] 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.
 - [6] 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.
 - [7] 林北征. 论生成式人工智能服务提供者的注意义务[J]. 法律适用, 2024(10): 148-163.