Open Journal of Legal Science 法学, 2026, 14(7), 263-269
Published Online July 2026 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2026.147220

Hans汉斯

生成式人工智能侵权归责原则的立场抉择与规范构造

——以全国首例AI“幻觉”侵权案的裁判逻辑为中心

林晶霞

宁波大学马克思主义学院，浙江 宁波

收稿日期：2026年6月18日；录用日期：2026年6月29日；发布日期：2026年7月29日

摘　要

生成式人工智能侵权归责原则的确定，是当前侵权法理论回应技术挑战的核心议题。全国首例因AI“幻觉”引发的侵权纠纷案，为这一争议提供了重要的司法实践样本。该案判决明确将生成式人工智能服务定性为“服务”而非“产品”，适用《中华人民共和国民法典》(后文简称《民法典》)第1165条第一款的一般过错责任原则，并从动态系统论出发构建了三层注意义务体系。然而，该判决所选择的归责方案与学界主张的过错推定说之间存在理论张力。本文以该案裁判逻辑为中心，系统检视生成式人工智能侵权的行为特性与归责困境，比较分析一般过错责任、无过错责任与过错推定责任三种方案的适用效果。研究认为，一般过错责任在举证分配层面存在结构性缺陷，过错责任则面临规范适用上的体系障碍，而过错推定责任在举证平衡、利益协调与制度激励方面具有比较优势。在此基础上，本文提出以过错推定为核心的规范构造方案，包括责任主体的类型化界分、举证责任的合理分配以及推翻推定过错的抗辩事由设计，以期在受害人救济与技术创新之间实现动态平衡。

关键词

侵权责任，归责原则，过错推定，注意义务，AI“幻觉”

Stance Selection and Normative Construction of Liability Principles for Generative AI Infringement

—Centered on the Judicial Logic of China’s First Case of AI “Hallucination” Infringement

Jingxia Lin

School of Marxism Studies, Ningbo University, Ningbo Zhejiang

文章引用: 林晶霞. 生成式人工智能侵权归责原则的立场抉择与规范构造[J]. 法学, 2026, 14(7): 263-269.
DOI: 10.12677/ojls.2026.147220

Received: June 18, 2026; accepted: June 29, 2026; published: July 29, 2026

Abstract

Determining liability principles for infringement caused by generative artificial intelligence is a central issue in current tort law theory's response to technological challenges. The first nationwide case involving an AI "hallucination" that triggered an infringement dispute provides a significant judicial precedent for this debate. The court ruling clearly classified generative AI services as "services" rather than "products", applied the general principle of fault liability under Article 1165, Paragraph 1 of the *Civil Code of the People's Republic of China* (hereinafter referred to as the *Civil Code*), and established a three-tier duty-of-care framework based on dynamic systems theory. However, the chosen liability approach in this judgment exhibits theoretical tension with the academic consensus favoring presumed fault liability. Centered on the reasoning of this case, this article systematically examines the behavioral characteristics and liability challenges associated with generative AI infringement, and comparatively analyzes the practical effects of three liability models: general fault liability, strict liability, and presumed fault liability. The study concludes that general fault liability suffers from structural flaws in burden allocation, while fault-based liability faces systemic barriers in legal application; in contrast, presumed fault liability offers comparative advantages in balancing evidentiary burdens, coordinating interests, and incentivizing innovation. Based on these findings, the article proposes a regulatory framework centered on presumed fault liability, including typological distinctions among liable parties, rational allocation of the burden of proof, and design of defenses to rebut presumed fault, aiming to achieve a dynamic balance between victim compensation and technological advancement.

Keywords

Liability for Infringement, Liability Principles, Presumption of Fault, Duty of Care, AI "Hallucination"

1. 问题的提出

2026 年 1 月 27 日，杭州互联网法院审结的全国首例因生成式人工智能模型“幻觉”引发的侵权纠纷案(以下简称“本案”)，引发了法学理论与实务界的广泛关注[1]。该案中，梁某于 2025 年 6 月 29 日使用某科技公司开发的生成式人工智能应用程序查询某高校报考信息，AI 模型生成了关于该高校存在某校区的不准确信息。梁某发现后进行纠正和指责，AI 仍继续回复称该校区确实存在，甚至“承诺”若内容有误将赔偿 10 万元，并建议梁某到杭州互联网法院起诉索赔。梁某遂以侵权为由起诉，要求科技公司赔偿损失 9999 元。法院经审理，驳回了原告的全部诉讼请求[2]。

这一判决之所以具有里程碑意义，不仅在于它是全国首例 AI“幻觉”侵权案，更在于它直面了生成式人工智能侵权责任认定中的一系列根本性问题：AI 能否独立作出意思表示？生成式人工智能致害应适用何种归责原则？服务提供者的注意义务边界何在？清华大学法学院教授程啸指出，该判决“对于今后

[1]https://zjsfgkw.gov.cn/art/2026/1/27/art_56_44943.html
[2]http://cyfy.stcourts.gov.cn/article/detail/2026/02/id/9205782.shtml

司法实践裁判人工智能侵权纠纷中准确适用侵权法的归责原则具有重要启示”[1]。

围绕生成式人工智能侵权归责原则，学界存在显著分歧。王利明教授主张适用一般过错责任，认为现行法“在整体上可以应对生成式人工智能带来的挑战”，不需要对现有规则进行大的调整[2]。吴太轩、邓朝辉则主张过错推定责任，认为一般过错和无过错责任“皆难完全满足现实的利益衡平要求”，可能造成“两造举证义务分配失衡、抑制产业发展积极性、消弭企业技术进步自觉性等贻害”，而过错推定则有望产生“内外两重适用优势”[3]。

归责原则的选择，不仅关乎受害人的救济效果，更深刻影响着生成式人工智能产业的发展方向与侵权法体系的适应性。本文以杭州互联网法院案的裁判逻辑为中心，在分析生成式人工智能侵权行为特性的基础上，比较不同归责方案的优劣，进而提出以过错推定为基点的规范构造方案。

2. 生成式人工智能侵权行为特性的法理审视

2.1. 作为“行为主体”而非“责任主体”的生成式人工智能

准确界定生成式人工智能的法律地位，是确定侵权归责原则的逻辑前提。杭州互联网法院在本案中明确指出：“在现行法中，享有民事权利，能够作出意思表示的民事主体仅有自然人、法人和非法人组织这三类。人工智能不具有民事主体资格，不能独立、自主作出意思表示。”[3]法院进一步指出，AI 生成的“赔偿承诺”不能视为服务提供者的意思表示，“某科技公司也没有通过将人工智能模型作为程序工具，设定或传达其意思表示的行为。”

然而，否定人工智能的民事主体地位，并不意味着否定其在事实层面的行为主体属性。吴太轩、邓朝辉概括了生成式人工智能侵权行为相较于传统侵权行为的四种独特属性：单独性——能够在无需外部协助的情况下单独完成侵权信息的创造性编撰并向用户直接发布；直接性——以自身“意志”直接作用于相对人权益；深度自主性——由于“智能涌现”效应，模型的输出结果开始呈现“不可理解、不可预测和不可控制”的趋势；纯技术性——其所谓的“自主”始终是人类智能的工具性延伸，无法改变其“物件”本性。

这四点特性共同指向一个核心结论：生成式人工智能是行为主体和事实主体，却非责任主体与法律主体。侵权行为的实施者与责任的最终承担者发生分离——这是确定归责原则时必须正视的结构性特征。

2.2. 传统侵权类型框架的解释困境

吴太轩、邓朝辉指出，将生成式人工智能侵权行为纳入传统侵权类型框架，面临多重解释障碍。生成式人工智能侵权行为既不同于单独侵权与数人侵权(因其行为主体并非法律主体)，也不同于物件致损的侵权(因其具有深度自主性)，还有别于自己责任与替代责任的侵权(因行为主体与责任主体分离但不存在监护、雇佣等特定关系) [3]。

这种“不可归类性”意味着，既有的侵权类型化方案难以直接为生成式人工智能侵权提供妥当的归责逻辑。归责原则的选择，不能简单套用某一传统侵权类型的既有规则，而必须回到侵权法的基本原理——过错责任原则及其例外——重新进行规范分析。

2.3. 侵权类型的进一步区分

从受害人权益类型出发，生成式人工智能侵权可进一步区分为不同类型，其归责逻辑也应有所差异：

人格权侵权。AI 生成诽谤信息、泄露隐私等情形，涉及对名誉权、隐私权等人格权的侵害。此类侵权中，受害人往往面临“损害难以量化”的困境。在我国法上，过错推定在人格权保护领域已有制度先

[3]同 2。

例(如《民法典》第1198条安全保障义务)，为过错推定在人格权侵权场景的适用提供了参照。

财产权侵权。AI提供错误投资建议、错误报考信息等致纯粹经济损失的情形。此类侵权的归责应更为审慎——纯粹经济损失的赔偿范围本身即受严格限制，一般过错责任可能在财产权侵权场景具有更强理由。

知识产权侵权。AI生成内容侵害著作权的情形。此类侵权的注意义务标准有所不同。本文认为“通知”规则在生成式人工智能作品侵权中具有类推适用的空间，服务提供者在接到权利人通知后未采取必要措施的，可据以认定过错。

区分不同类型后，“过错推定”主张可能需要进一步精细化——是否所有类型均适用过错推定，还是仅针对人格权侵害等特定类型？下文将基于此区分展开进一步分析。

3. 归责原则的立场抉择：三种方案的比较分析

3.1. 一般过错责任：司法选择的逻辑与局限

杭州互联网法院案的核心裁判贡献之一，是明确了对生成式人工智能侵权应适用一般过错责任原则。法院给出了四点理由：其一，生成式人工智能服务“缺乏具体、特定的用途与合理可行的质检标准”，不具备“产品”的构成要件；其二，其生成的信息内容本身“通常不具备民法典侵权责任编所指的高度危险性”；其三，服务提供者“缺乏对生成信息内容足够的预见和控制能力”；其四，适用无过错责任原则“可能会不当加重服务提供者的责任，限制人工智能产业的发展”[4]。

王利明教授进一步从法理层面论证了这一立场：其一，从民法典侵权责任编对无过错责任原则的规定来看，无过错责任以侵权行为“对他人的人身、财产安全具有较高的危险性”为前提，而生成式人工智能“充其量只是向人类提供信息”，通常不危及人身或财产安全；其二，从政策导向看，对服务提供者采取无过错责任原则“可能会不当加重其责任，从而限制AI产业的发展”[4]。

然而，一般过错责任原则在适用中面临一个现实难题：举证分配的不平衡。生成式人工智能的运行机制具有高度的封闭性和技术复杂性，用户难以获取模型训练数据、算法设计、输出机制等内部信息。技术方握有信息优势，若继续沿用一般过错责任原则，权利人必须证明对方过错，技术壁垒使举证难以实现，可能实质上剥夺受害人的救济机会。

具体而言，在AI生成诽谤信息损害他人名誉权的场合，受害人若要证明服务提供者“应当知道”该诽谤信息的存在并阻止其生成，就需要了解该AI模型的训练数据是否包含该主体的相关信息、过滤机制为何未能识别该诽谤信息、相似内容在以往的生成记录中是否曾被标记——而这些信息全部掌握在服务提供者手中，受害人根本无法获取。若举证责任不向服务提供者转移，受害人在诉讼中除了证明AI确实生成了诽谤信息之外，几乎没有能力提供任何关于过错的实质性证据。这正是过错推定制度所要解决的核心问题。

3.2. 无过错责任：审慎否定的法理分析

关于生成式人工智能侵权应否适用无过错责任(产品责任)，杭州互联网法院明确予以否定，其核心理由在于：依据《生成式人工智能服务管理暂行办法》[5]，生成式人工智能服务明确定位为“服务”范畴。传统产品的物理形态、功能边界和风险范围是明确且相对固定的，而生成式人工智能生成的内容“是算法基于海量数据、对用户开放式指令的即时响应，其每一次输出都是不可完全复现的独特生成”[5]。

从民法典侵权责任编的规定来看，无过错责任以侵权行为“对他人的人身、财产安全具有较高的危

[4]https://sfb.pkulaw.com/qikan/077351e9f7815fdabd567d4d13d7464cbdfb.html
[5]https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm

险性”为前提，而生成式人工智能通常不满足这一前提。

在政策层面，在技术发展初期对其苛以过重的严格责任，可能产生“寒蝉效应”，导致某些创新技术无法落地应用。因此，无过错责任方案难以获得正当性支持。

3.3. 过错推定：理论优势与比较法启示

在一般过错责任与无过错责任之间，过错推定责任提供了一条中间道路。从理论基础和实践效益两个维度检视后发现：过错推定有望产生内外两重适用优势。

“对内优势”：将举证责任转移至服务提供者，能够有效克服用户在技术信息层面的举证障碍，实现举证分配的实质公平。服务提供者作为技术活动的组织者和控制者，最有能力证明自己是否已尽到合理注意义务，由其对“无过错”承担举证责任符合“证据距离”原则。

“对外优势”：在保障受害人救济机会的同时，服务提供者仍可以通过证明自己已尽到合理注意义务来推翻过错推定，避免无过错责任对产业的过度压制。这种设计既能激励服务提供者积极采取防范措施，又不至于使其沦为“信息风险的保险人”。

比较法上，欧盟《人工智能责任指令》(Proposal for a Directive on adapting non-contractual civil liability rules to artificial intelligence)提案为过错推定责任提供了重要参照。该指令引入“因果关联之推定”：如果相关过错被证明，且该过错与人工智能效能之间产生因果关联具有合理可能，受害人无需详释损害如何由过错引起。同时引入“获取证据之权利”：当高风险 AI 系统疑似引起损害，法院可下令相关系统供应者披露证据[6]。欧盟 2024 年通过的修订版《产品责任指令》(Directive (EU) 2024/2853)更进一步将人工智能系统纳入产品责任范畴，在特定情形下实行举证责任倒置[7]。这一制度设计与中国学界主张的过错推定具有内在契合性——均在受害人举证困难与技术复杂性之间寻求平衡。

关于过错推定的教义学基础，徐伟教授从保护性规范理论出发进行了论证。他指出，服务提供者注意义务的来源有二，即“法令上义务”和“一般防范损害发生之义务”[6]。我国现行法中涉服务提供者的保护性规范提供了丰富的注意义务内容。在个案纠纷落入保护性规范所保护的“人的范围”和“物的范围”的前提下，违反保护性规范的可推定服务提供者存在过错。这为过错推定提供了教义学上的制度接口。

4. 过错推定责任的规范构造

4.1. 责任主体的类型化界分

过错推定责任的规范构造，首先需要明确责任主体的范围与类型，应将生成式人工智能服务提供者确定为责任主体，并进一步界分为开发者和运营者，要求二者各自就其过错承担独立举证责任。

这一界分的合理性在于：开发者处于 AI 模型的起点，负责模型搭建、训练数据选取与算法设计，应在训练阶段承担源头管控义务；运营者面向用户提供服务，负责内容输出、风险提示与投诉处置，应在服务阶段承担过程管控义务。二者在注意义务的内容与履行方式上存在差异，分别举证有利于准确认定过错。

为增强该界分的可操作性，需对“开发者”与“运营者”提供更为清晰的法律界定标准。本文建议，“开发者”可界定为对人工智能模型进行初始设计、训练数据选择、算法架构搭建及模型版本迭代负有实质性控制能力的组织或个人；“运营者”则可界定为面向终端用户部署、提供人工智能服务，并有权对服务内容、交互方式及用户管理进行日常控制和调整的组织或个人。在多主体参与的复杂供应链，如

[6]http://m.legalweekly.cn/whlh/2022-10/20/content_8792150.html
[7]https://www.dwcon.cn/post/4130

采用模型由 A 公司开发、B 公司进行微调、C 公司提供云基础设施、D 公司面向用户运营的模式，责任分配不宜采取单一主体模式，而应根据各主体对致害环节的实际控制力与防范能力进行分层认定。具体而言，若损害源于基础模型的先天缺陷(如训练数据中的系统性偏见)，开发者应承担主要责任；若损害源于微调阶段引入的特定风险或运营阶段的服务管理缺失，则相应由微调者或运营者承担责任。各主体可依据《民法典》第 1168 条关于共同侵权的规定，或第 1171 条、第 1172 条关于无意思联络数人侵权的规定，承担连带责任或按份责任。

杭州互联网法院案确立的三层注意义务体系，为过错推定中的“过错”认定提供了可操作的判断标准：一是对“有毒”、有害、违法信息负有严格审查义务(结果性义务)；二是需以显著方式提示 AI 生成内容可能不准确的固有局限性(方式性义务)；三是应尽功能可靠性的基本注意义务，采取同行业通行技术措施(如检索增强生成技术)提高生成内容准确性(方式性义务)。

这一动态化的注意义务框架，与过错推定责任的运作逻辑高度兼容：服务提供者若无法证明其已履行上述注意义务，即可推定其存在过错；反之，若能证明已采取合理措施，则可推翻过错推定。

4.2. 举证责任的合理分配

在过错推定框架下，举证责任的分配遵循如下逻辑：受害人仅需证明损害事实、加害行为以及二者之间的因果关系(适用相当因果关系标准)，而无需证明服务提供者的过错；服务提供者若欲免责，须举证证明其不存在过错——即已尽到合理注意义务。

这一分配方案回应了生成式人工智能侵权中“受害人举证难”的核心困境。过错推定将举证责任转移至技术信息的持有方，有助于平衡双方在诉讼中的举证能力。

关于过错推定的法律依据，徐伟教授从保护性规范理论出发，指出现有人工智能公法规范中包含的保护性规范，可转介为侵权法中的注意义务内容，违反此类公法规范的行为可触发过错推定的适用[7]。在个案纠纷落入保护性规范所保护的“人的范围”和“物的范围”的前提下，违反保护性规范的可推定服务提供者存在过错。这为过错推定责任提供了教义学上的制度接口。

4.3. 推翻推定过错的抗辩事由设计

为防止过错推定沦为变相的严格责任，须为服务提供者设计合理的抗辩事由。结合杭州互联网法院案的裁判经验与学理讨论，以下抗辩事由可成为服务提供者推翻过错推定的有效路径：

其一，已尽显著提示义务。服务提供者在应用程序欢迎页、用户协议及交互界面的显著位置呈现 AI 生成内容功能局限的提醒标识，并确保提示方式的“显著性”。在涉及重大利益的特定场景(如升学报考、医疗咨询等)进行了正面、即时的“警示提醒”。本案中，法院即以此作为认定被告已尽注意义务的依据之一。

其二，已采取行业通行技术措施。服务提供者已采用检索增强生成(RAG)等现行行业通行的技术措施以提升生成内容的准确性，且对内容准确性负方式性义务而非结果性义务——不能因生成内容不准确就当然认定其存在过错。本案中，被告的大模型已完成国家备案和安全评估，并采用了 RAG 技术，法院据此认定其尽到了功能可靠性的基本注意义务。

其三，损害超出可预见与控制范围。服务提供者对生成内容缺乏足够的预见和控制能力，且损害的发生与用户自身的决策行为存在显著关联。本案中，法院以相当因果关系标准审查后认定，AI 生成的不准确信息“并未实质性地介入或影响原告的报考决策过程”，二者不存在因果关系。

为将“可预见性”“技术可行性”等模糊概念转化为具有操作性的判断指引，可参考以下标准：在“可预见性”的判断上，应考察损害类型是否属于该领域一般专业人士在模型开发或运营时能够预见的

风险范围，具体可参考同行业类似技术水平模型在相同或相似使用场景下的已知风险清单；在“技术可行性”的判断上，应以行业通行的技术标准为主要依据，参考国家标准化管理委员会或相关行业协会发布的技术规范与最佳实践指南，同时兼顾中小型服务提供者在资源投入方面的合理差异，但不得以规模为由完全免除其应尽的基本注意义务。

其四，遵循保护性规范。服务提供者能够证明其行为符合现行法律、行政法规中对生成式人工智能服务的公法要求，且该规范的遵守与损害防范具有实质关联性。

5. 结语

全国首例 AI“幻觉”侵权案的判决，为生成式人工智能侵权责任的司法认定提供了重要的先导性指引。该案选择一般过错责任原则并构建动态注意义务体系，在个案中实现了权益保护与技术创新的平衡。然而，从制度完善的角度看，一般过错责任在举证分配层面存在结构性缺陷，可能影响受害人救济的实效性。

过错推定责任既能克服举证障碍，又为服务提供者保留了抗辩空间，在现行法框架内具有适配性——其教义学基础可从《民法典》第 1165 条第二款 [8]的法律推定和保护性规范的侵权法转介两条路径获得支撑。欧盟《人工智能责任指令》[9]引入的“因果关联之推定”与“获取证据之权利”两项制度设计，为过错推定责任提供了比较法上的参照[8]。本文主张，应通过司法解释或指导性案例，逐步确立以过错推定为核心的生成式人工智能侵权归责方案，并在实践中不断细化注意义务标准，以实现技术进步与法治保障的动态平衡。

参考文献

[1] 程啸. 司法实践为 AI 治理提供法理支撑[N]. 法治日报·法治周末, 2026-01-26(3).

[2] 王利明, 包丁裕睿. 论“通知”规则在生成式人工智能作品侵权中的类推适用[J]. 比较法研究, 2025(4): 1-13.

[3] 吴太轩, 邓朝辉. 生成式人工智能侵权归责原则的比选与适用[J]. 电子知识产权, 2025(8): 4-18.

[4] 王利明. 生成式人工智能侵权的归责原则与过错认定[J]. 法学, 2025(4): 15-30.

[5] Rodríguez de Las Heras Ballell, T. (2025) Mapping Generative AI Rules and Liability Scenarios in the AI Act, and in the Proposed EU Liability Rules for AI Liability. *Cambridge Forum on AI: Law and Governance*, **1**, e5. https://doi.org/10.1017/cfl.2024.8

[6] Pehlivan, C.N., Forgó, N. and Valcke, P. (2025) AI Governance and Liability in Europe: A Primer. Kluwer Law International.

[7] 徐伟. 生成式人工智能服务提供者侵权过错的认定[J]. 法学, 2024(7): 110-124.

[8] 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.

[8]https://zqdzfy.gov.cn/uploads/ueditor/file/20250717/1752715098194576.pdf
[9]https://cclp.sjtu.edu.cn/Show.aspx?info_lb=672&info_id=5814&flag=648