

生成式人工智能网络谣言法律责任研究

张荣金

重庆工商大学法学与社会学学院, 重庆

收稿日期: 2026年3月20日; 录用日期: 2026年5月14日; 发布日期: 2026年5月22日

摘要

随着以ChatGPT、Sora为代表的生成式人工智能技术的爆发式发展, 网络谣言的生成与传播机制发生了质变。生成式人工智能技术不仅极大地降低了网络谣言的制造成本, 更通过深度伪造等方式使网络谣言内容呈现前所未有的高度逼真性, 借助算法推荐和社交机器人实现精准化、智能化的裂变式传播, 构成了严峻挑战。现行法律规制在应对此类谣言时面临诸多困境, 尤其是在责任主体认定和归责原则适用上存在模糊地带。当谣言的直接生产者是非人时, 如何在技术开发者、服务提供者 and 使用者之间分配责任成为难题。基于“控制力”理论, 本文认为, 应重构一种以控制力为核心的多元、分层责任体系。在归责原则上, 应以过错责任为基础, 并针对不同主体和场景引入过错推定和严格责任作为补充。通过明确各方主体的注意义务, 并类推适用“通知-必要措施”规则, 可以在鼓励技术创新的同时, 有效遏制网络谣言的危害, 实现技术发展与社会安全的协调发展。

关键词

生成式人工智能, 网络谣言, 法律责任, 归责原则

Research on Legal Liability of Network Rumors Generated by Artificial Intelligence Networks

Rongjin Zhang

School of Law and Sociology, Chongqing Technology and Business University, Chongqing

Received: March 20, 2026; accepted: May 14, 2026; published: May 22, 2026

Abstract

With the explosive development of generative AI technologies represented by ChatGPT and Sora, the mechanisms of rumor generation and dissemination have undergone fundamental transformations.

These technologies not only drastically reduce the cost of creating online misinformation but also enable unprecedentedly realistic content through deepfake techniques. Leveraging algorithmic recommendations and social bots, they facilitate targeted, intelligent viral propagation, posing significant challenges to legal frameworks. Current regulatory systems face multiple difficulties in combating such rumors, particularly regarding ambiguous liability determinations and attribution principles. When algorithms rather than humans directly generate misinformation, allocating responsibility among tech developers, service providers, and users becomes particularly complex. Grounded in the “control theory”, this study proposes reconstructing a multi-tiered liability framework centered on control capacity. The attribution principle should adopt fault-based liability, supplemented by presumed fault and strict liability mechanisms tailored to different entities and scenarios. By clarifying each party’s duty of care and applying the “notice-necessity” rule through analogy, we can effectively curb the harm caused by online rumors while encouraging technological innovation, achieving harmonious development between technological progress and social security.

Keywords

Generative Artificial Intelligence, Online Rumors, Legal Liability, Principle of Attribution

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

谣言作为一种古老的社会文化现象，与人类社会相伴相生，其生成、传播与所处时期的媒介技术紧密相连。从口耳相传到印刷媒介，再到网络平台，每一次媒介革命都重塑了谣言[1]。近年来，以 ChatGPT、Sora 等为代表的生成式人工智能技术迎来了爆发式发展，人类社会迈入一个全新的智能时代。截至 2025 年 12 月，我国生成式人工智能用户规模已达 6.02 亿人[2]。技术在解放生产力、驱动创新的同时，也带来了前所未有的挑战。生成式人工智能的出现，极大地改变了网络谣言的生态。

在以往，制作一条足以乱真的虚假视频或图片需要专业的技术和大量时间，而如今，借助生成式人工智能工具，普通用户仅需简单的指令就能生成高度逼真的文本、图像、音频乃至视频内容。例如，利用生成式人工智能技术生成的“云南山体滑坡灾害已致 8 人遇难”的虚假新闻，其语言精练、内容全面，足以误导大量网民。这种低成本、高效率的造谣模式，使得谣言的生产可能成为一种普遍的社会现象，极大地增加了监管的难度[3]。

面对生成式人工智能技术赋能下的网络谣言，现有的法律责任框架遭遇了挑战。责任主体应如何认定？是仅追究提供提示词的用户？还是提供生成式服务的平台，抑或是训练大模型的技术开发者？归责原则应如何适用？是适用产品责任的无过错原则，还是网络服务提供者的过错责任原则？这些问题不仅是理论上的思辨，更是亟待解决的实践难题。

2. 生成式人工智能网络谣言的特性及危害

2.1. 生成式人工智能网络谣言的特征

在生成式人工智能技术的加持下，网络谣言的生成与传播呈现出若干显著的新特征，使其治理变得愈发复杂和困难。

2.1.1. 生成机制

传统谣言的生成高度依赖个体对信息的搜索、筛选与拼凑，需耗费大量时间精力。在生成式人工智能技术的介入下，谣言的制造变得规模化，造谣者能够以极低的成本批量生产谣言，造谣者仅需提供关键词或简要指令，生成式人工智能工具即可自动抓取网络信息，融合热点要素与随机素材，在极短时间内完成内容的批量生产[4]。在2024年，江西警方披露一起利用人工智能造谣的案例，江西南昌一家MCN机构实际控制人王某某通过AI软件生成内容的方式，这家MCN机构短时间内生成了大量文章，最高峰一天能生成4000至7000篇，日均生成虚假文章可达四千至七千篇，其内容看似信息详实，实则完全虚构¹。这种生成机制的质变，标志着谣言生产已进入低门槛、高效率的阶段。随之而来的，是行为主体与责任来源的模糊化，当侵权内容由人类恶意与机器创作共同完成时，责任主体的界定便成为首要的法律难题。

2.1.2. 内容形态

传统谣言多依赖现有素材的断章取义或生硬拼接，通过对现有素材断章取义、张冠李戴，将不同来源的文字、图片生硬组合。此类谣言因素材来源杂、拼接痕迹重而易于识别。而生成式人工智能能够从海量数据中学习，能依据简单指令生成高度逼真的内容²。有研究通过对典型AI谣言案例进行语义网络分析，发现其文本往往结构复杂、语义离散度高，真伪辨识难度极大[5]。以“首例智能驾驶致死案宣判”这则谣言为例，从事故认定到诉讼细节再到法院判决结果，整篇文章在逻辑上和专业性上都难以察觉出漏洞，致使多家媒体一度误报³。此异化从根本上动摇了以应知、明知为核心的传统网络平台注意义务，传统模式下，平台对明显虚假或侵权内容负有审查责任。但生成式人工智能网络谣言在形式上高度逼真，依靠常规审核手段难以实现有效辨识。若仍要求平台对每条内容进行事实性审核，无异于施加了超出其能力范围的严格责任。

2.1.3. 传播模式

生成式人工智能网络谣言的传播机制，其核心特征在于算法分发的介入。与传统依赖编辑筛选或社交关系链相比，基于用户画像与内容特征的算法推荐，能够实现谣言与受众的精准匹配，从而在极短时间内完成高效、广泛的扩散[6]。有研究通过对近年35起高热度的AI谣言案例进行定性比较分析发现，AI谣言传播效果的关键因素，已不再主要依赖于信息源的权威性，而更侧重于内容本身的呈现质量与受众的互动行为[7]。这一转变对以“通知-删除”规则为基石的平台责任框架构成了结构性挑战。当平台通过算法主动排序、推荐并放大谣言时，其角色已从被动的角色转变为主动角色。算法行为本身成为影响损害范围与程度的关键变量。在算法助推下，谣言的危害可能呈爆发式形成，事后的删除往往难以挽回社会影响。

2.2. 生成式人工智能网络谣言的危害

生成式人工智能网络谣言对社会秩序和公民权益构成了前所未有的危害，其危害性体现在多个层面。

首先，严重侵蚀社会信任体系。当深度伪造技术可以轻易地伪造任何人的言行时，公众对信息真实性的判断能力被极大削弱，“眼见为实”的传统认知被颠覆。虚假信息的泛滥不仅会降低公信力，更会引发普遍的社会信任危机，导致社会共识难以形成。在信息真假难辨的环境中，人们可能会对各类信息

¹参见《一天写7000篇！MCN机构利用AI造谣，细节曝光》，载京报网，<https://news.bjd.com.cn/2024/06/14/10804183.shtml>，2025年11月20日访问。

²参见《AI伦理观察 | 从“截洪流”到“堵水源”：AI生成谣言的治理升级》，载新华网，<https://www.xinhuanet.com/finance/20250806/b03c64200c2746c6a9b63af6c496d26e/c.html>，2026年2月20日访问。

³参见《AI谣言舆情特征及风险研判》，载法治网，http://www.legaldaily.com.cn/index/content/2025-05/16/content_9183906.html，2026年2月20日访问。

源产生广泛质疑，引发信任危机[8]。

其次，严重侵害公民个人合法权益。利用技术制作和传播他人不雅照片、利用换脸技术进行电信诈骗、编造虚假信息诽谤他人名誉等行为，直接侵害了公民的隐私权、肖像权、名誉权和财产权。这些行为不仅给当事人带来巨大的精神痛苦和财产损失，还可能使得公民的个人信息被滥用，导致公民的财产安全遭受损害[9]。再次，严重扰乱社会公共秩序和经济秩序。在突发公共事件(如自然灾害)期间，涉灾涉险的谣言极易引发社会恐慌，导致物资抢购、交通混乱等问题，严重干扰正常的应急管理工作。

3. 生成式人工智能网络谣言的责任困境

尽管我国已初步形成了涵盖《网络安全法》《民法典》《刑法》及一系列部门规章的网络信息内容治理法律体系，但在应对生成式人工智能带来的新型谣言挑战时，这些规定暴露出了明显的局限性和不适应性。技术的发展速度远远超过了法律的更新速度，导致法律规范在面对生成式人工智能网络谣言时，无论是在责任主体的认定、归责原则的适用，还是在因果关系的证明上，都面临着严峻的困境。

3.1. 责任主体认定困境

传统网络侵权责任的认定，通常围绕内容发布者和网络服务提供者平台展开。然而，生成式人工智能网络谣言的生成与传播表现为使用者提示、生成平台响应、传播平台扩散的路径，涉及使用者、内容生成平台、内容传播平台三类核心主体，以及作为技术工具的生成式人工智能本身。这种多元主体的参与，使得传统的“谁发布，谁负责”原则难以简单适用。

关于生成式人工智能是否具备法律主体地位，是否能独立承担法律责任，当前学界虽然存在“拟制说”“电子代理人说”“电子人格说”等肯定说，但主流意见仍坚持其工具属性，认为当前生成式人工智能仍属于人类创作工具，其不具备独立的意志与责任能力[10]。因此，在法律层面，无法将生成式人工智能作为独立的责任主体追究。

然后对于生成式人工智能的使用者即造谣者，其作为谣言的发起者，是生成式人工智能网络谣言的应然主体，必然承担源头责任。因而生成式人工智能网络谣言的侵权责任困境主要是体现在生成式人工智能服务提供者上。在生成式人工智能网络谣言的链条中，服务提供者是风险的控制中枢、技术的源头和扩散的放大器。它是法律必须规制的核心，也是《生成式人工智能服务管理暂行办法》等法规的直接规制对象。其作为服务的提供方，从用户的活跃中获益，似乎应承担更重的注意义务。但平台往往辩称自己仅是“技术中立”的工具提供者，生成何种内容取决于用户，平台难以对海量、瞬时生成的内容进行事前审查。这种观点在传统网络服务中尚有一定市场，但在平台深度参与内容生成与分发的时代，其合理备受质疑。

3.2. 归责原则困境

基于侵权法原理和我国现行法律，在无特别规定的情况下，如果能够证明损害是因生成式人工智能模型的选择问题造成的，则应当由其研发者承担过错责任；如果损害是因生成式人工智能操作者的操作不当引起的，则应当由其操作者承担过错责任，即对生成式人工智能侵权也应当适用过错责任原则[11]。但是在生成式人工智能网络谣言场景下，传统的过错责任原则面临显著适用障碍。

其一，生成式人工智能技术特性与过错责任的主观归责基础不相契合。过错责任原则以行为人的主观可归责性为前提，强调其故意或过失所体现的不正当性。然而，生成式人工智能输出内容并非由服务提供者直接意志控制，而是由使用者控制，再通过数据输出而成。服务提供者可通过人工标注等方式干预生成内容，却难以全面预防网络谣言的产生。在此情形下，服务提供者既无主观制造谣言的意图，也难以被认定为存在重大过失，导致其主观过错难以成立。

其二，即便对过错进行客观化解读，以违反注意义务作为认定标准，受害人在实践中仍面临举证困境。注意义务的认定通常依据行业普遍技术水平，而生成式人工智能尚在快速发展阶段，缺乏稳定的行业标准可供参照。受害人需证明服务提供者未采取合理措施防止谣言生成，但相关技术决策、内部审核机制及模型训练数据往往处于黑箱状态，外界难以获取证据。此外，生成式人工智能内容生成具有随机性和不可完全预测性，进一步加大了证明服务提供者存在义务违反的难度[12]。

3.3. 责任边界困境

责任边界困境首先体现为服务提供者与恶意使用者之间的内部责任划分模糊。《生成式人工智能服务管理暂行办法》确立了提供者、使用者的主体框架，并且规定了提供者对违法内容的处置义务，但对于用户通过精心设计的恶意提示词刻意绕过过滤机制的行为，服务提供者应承担何种程度的识别与阻断义务，法律并未明确。更复杂的困境在于损害发生过程中的动态化、阶段化责任分配。有学者提出场景化思路，将责任对应到训练、生成、传播、移除等不同阶段[13]。例如，训练阶段的数据偏见导致模型易在某些领域造谣，责任重心在开发者；生成阶段，服务提供者通过实时安全过滤器的效力，直接影响有害内容是否被输出，此时其运营责任凸显；在传播阶段，若平台在收到确切举报后未及时采取措施阻止谣言扩散，则可能承担扩大损失的责任。然而，这种理想划分面临现实挑战：一是证据固化的困难，用户提示词与生成式人工智能生成响应瞬息万变，难以完整记录还原；二是多阶段因素竞合，一个谣言可能同时源于训练数据缺陷、实时过滤疏漏和传播延迟，各原因力比例难以量化；三是技术措施的合理边界，过度过滤会损害工具实用性，引发用户不满，而过滤不足则面临法律风险，服务提供者在这期间的裁量权如何被司法评价，同样缺乏标准。

4. 生成式人工智能网络谣言域外经验

4.1. 欧盟经验

与我国针对生成式人工智能开展专门立法的路径不同，欧盟采取的是制定一般法意义上的综合性“人工智能法”模式。自2021年4月欧盟委员会公布《人工智能法案》提案以来，历经2023年5月“修正案草案”以及2024年3月欧洲议会“立法决议”等多轮修订，该法最终于2024年8月1日正式生效。欧盟《人工智能法》并未对生成式人工智能进行单独立法规制，而是通过法律解释与定义，将其明确归属于通用人工智能模型及其规制范畴。此种模式的显著优势在于能够维持法律体系的相对稳定性：即便未来新型人工智能技术出现，立法者亦可依托既有框架中的根本理念、基本原则及包容性法律定义进行规范，从而避免因技术快速迭代导致的法律漏洞，为后续法律修订争取宝贵的应变时间。

在责任分配机制上，欧盟构建了一套覆盖全过程的责任体系。具体而言，该法明确了人工智能系统提供者作为风险源头承担主要的合规与治理义务；部署者则负责应用阶段的持续监督与必要的人工干预；同时，中间环节主体亦负有相应的合规审查与信息传递义务。对我国而言，欧盟经验具有双重借鉴价值：其一，我国现行《生成式人工智能服务管理暂行办法》作为部门规章，存在法律效力位阶偏低、体系性整合不足的局限，适时制定法律层级的人工智能法有助于构建我国人工智能规制框架；其二，欧盟所确立的责任分配机制，可为厘清我国实践中生成式人工智能服务提供者、技术支持者与使用者之间复杂的责任边界提供重要的制度参考[14]。

4.2. 美国经验

美国作为ChatGPT等生成式AI应用的发源地与产业聚集地，在技术应用层面拥有显著优势。然而，在监管策略上，美国并不同于欧盟的集中统一立法路径。在联邦层面，白宫发布了一系列行政命令与指

导意见，如《关于安全、可靠、值得信赖地开发和和使用人工智能的行政命令》⁴，试图通过政策引导强化治理；而在州层面，以加利福尼亚州为代表的部分地区积极探索。

在内容治理的核心依据方面，美国长期依赖 1996 年《通信规范法》第 230 条的规定。该条款确立了安全港原则，即网络服务提供者对于用户发布的第三方内容一般不承担法律责任，同时赋予其自主进行内容审核的广泛权限。这一免责条款为互联网平台提供了巨大的自治空间。然而，随着生成式 AI 技术的崛起，第 230 条的适用边界正面临严峻挑战。核心争议在于：传统上第 230 条旨在保护交互式计算机服务对第三方内容的中立传输，但生成式 AI 具备高度的内容创造性与自主性，其输出的内容在本质上并非单纯对用户指令的机械反映，这使得 AI 服务提供者极有可能构成法律意义上的信息内容提供者而不再是中立平台。若此定性成立，AI 生成的有害或虚假内容将无法再依据第 230 条获得豁免，平台将面临直接的侵权或违法责任风险^[15]。

5. 生成式人工智能网络谣言责任体系的完善路径

法律规制的目标应是在鼓励技术创新与防范社会风险之间寻求平衡，这要求我们必须在现有法律框架的基础上，结合生成式人工智能网络谣言的技术特性，对责任体系进行重构与完善。

5.1. 引入控制力理论

当行为主体对自身行为拥有足够的支配能力时，该行为才能被认定为是行为主体的行为，进而可以归责；一个人原则上不应为不受自己控制的行为承担责任^[16]。而如果某一主体无法控制自己的行为，而另一主体又基于特定关系对该主体享有控制力时，控制方应当承担法律责任。这一法理在我国现行民事立法中亦有具体的制度体现。《中华人民共和国民法典》第 1188 条规定的监护人责任、第 1191 条规定的用人单位责任等，都是责任主体基于特定的法律关系或事实状态，对直接行为人的行为拥有支配力或重大影响力，故应对该行为造成的损害承担责任^[17]。换言之，当某一主体对另一主体的行为拥有控制力时，法律便将该行为的效果归属于控制方；控制力的有无与强弱，直接决定了责任的归属与轻重。

将上述逻辑延伸至生成式人工智能网络谣言的场景，可以发现各方主体与 AI 系统之间存在着类似的控制与被控制关系。生成式人工智能本身尚不具备法律主体资格，但不同主体对谣言的生成与传播，拥有不同程度的实际控制力。技术开发者通过对模型架构的设计、训练数据的筛选以及模型对齐策略的调校，影响生成式人工智能的输出倾向。服务提供者则通过设置交互规则、部署安全过滤机制、调控算法分发逻辑等方式，对生成式人工智能的运行施加着持续而有效的控制。而使用者通过输入提示词与具体指令，对生成式人工智能网络谣言的最终生成起到了终端控制的作用。尽管 AI 并非法律意义上的“人”，但各方主体对其运行过程的支配关系，与替代责任中控制方对被控制方的支配关系具有结构上的同构性。因此，将生成式人工智能网络谣言各主体的责任与其实际控制力相挂钩，便与替代责任背后的控制理论形成了内在的逻辑同构。

因而控制力可以作为生成式人工智能网络谣言责任分配的工具。控制力越强的主体，可期待的注意义务标准越高，责任的边界也越宽。下文将从主体责任、归责原则与通知规则三个层面，提出具体的法律完善建议。

5.2. 厘清网络谣言主体责任

在生成式人工智能网络谣言的生成与传播链条中，不同主体对谣言的生成到扩散，拥有不同程度的实际控制力。因此可以将各主体的控制力进一步提炼为技术控制维度、过程控制维度与终端控制维度三

⁴<https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>.

个层面，以此作为界定其注意义务层级与责任边界的基本标尺。

第一，强化技术开发者与模型训练者的源头责任。作为技术控制维度的控制者，开发者和训练者对模型的底层架构、训练数据和算法逻辑拥有最强的控制力，理应承担最高的注意义务。其控制力来源于对模型底层架构的设计、训练数据的筛选以及模型对齐策略的调校，从根本上影响着生成式人工智能系统“能生成什么”和“倾向于生成什么”。应明确其在数据处理环节的事实性审查义务，要求其在训练语料中剔除已知的谣言库数据，并通过技术手段，引导模型在面对缺乏事实依据的提示词时，输出“拒绝回答”或“提示不确定性”的结果，从源头上压缩提示词转化为网络谣言的空间。

第二，明确服务平台的责任。平台是作为连接技术与用户的桥梁，是过程控制维度的核心主体，也是发现和治理虚假信息的关键节点。针对生成式人工智能网络谣言高度逼真的特点，平台必须严格落实《生成式人工智能服务管理暂行办法》中的标识义务，通过添加显式水印和隐式数字水印，明确提示受众“该内容由 AI 生成”，从而打破谣言的“逼真性”伪装，降低公众的误信率。同时，平台应建立与权威媒体、公安机关联动的动态辟谣机制，一旦发现生成式人工智能生成的虚假信息在平台内传播，应立即采取阻断措施^[18]。

第三，界定使用者的恶意造谣责任。使用者是谣言生成的直接指令者，处于终端控制维度的核心位置。对于明知或应知内容虚假，仍通过恶意提示词诱导生成式人工智能生成谣言，并将其发布至公共网络空间以博取流量或损害他人利益的用户，应依法追究其直接的民事侵权责任；情节严重扰乱公共秩序的，应依法追究行政乃至刑事责任。

以前文述及的南昌 MCN 机构利用 AI 批量造谣案为例，在该案中，王某某通过输入恶意提示词，指令 AI 工具生成数千篇虚假社会新闻，其行为处于终端控制维度的核心位置，是损害发生的直接触发者，应承担直接责任。服务平台则处于过程控制维度，若其未能部署针对社会新闻类敏感内容的安全过滤机制，则构成对过程控制义务的违反，应承担相应的平台责任。而 AI 模型的开发者处于技术控制维度，若训练数据中混杂了大量不可靠新闻源，或模型对齐策略未能有效抑制虚假社会新闻的生成倾向，则开发者亦应承担源头治理层面的相应责任。

5.3. 重构网络谣言归责原则

在明确各主体基于控制力的责任边界之后，还需在归责原则层面进行相应的制度设计，以保障责任分配的公正性与可操作性。针对生成式人工智能网络谣言侵权的复杂性，单一的归责原则已难以适应。未来的法律框架可构建一个多元化、分层级的归责体系，根据不同侵权场景的风险程度和归责需要，灵活选择最恰当的责任形式。

首先，坚持以过错责任为一般原则，合理界定对生成式人工智能网络谣言生成及传播的注意义务。对于生成式人工智能技术服务提供者而言，要求其对所有生成内容的绝对真实性承担责任是不现实的，这违背了当前大语言模型存在幻觉的技术客观规律。因此，过错的认定不应以是否生成了虚假信息为唯一标准，而应审查服务提供者是否尽到了合理防范谣言的注意义务。具体而言，这种注意义务应包括：是否在模型训练阶段引入了高质量的事实核查数据集；是否在生成阶段部署了针对敏感领域(如涉政、涉灾)的关键词拦截机制；是否为用户提供了便捷的辟谣和投诉渠道。只有当平台怠于履行上述防范虚假信息的合理义务时，才认定其存在过错。

其次，引入过错推定责任，规制算法推荐对谣言的助推行为。当平台不仅提供内容生成服务，还通过内置的算法推荐机制主动将生成的虚假信息分发、推送到公共场域，从而引发谣言的裂变式传播时，其角色已从工具提供者转变为内容分发者。在此场景下，考虑到受害者难以获取平台算法分发逻辑的证据，应适用过错推定原则。即一旦证明平台算法主动推荐了严重失实的谣言，即推定平台未尽到审慎核

查义务,由平台承担证明其算法机制不存在缺陷的举证责任[19]。

5.4. 重构通知规则

传统的通知-删除规则在应对 AIGC 网络谣言时显得滞后且单一。生成式人工智能网络谣言的传播具有瞬时爆发性,事后删除往往于事无补。因此,必须结合谣言治理的需求,对《民法典》中的“通知-必要措施”规则进行重构。

首先,针对生成式人工智能网络谣言,平台的“必要措施”不应仅局限于“删除”。对于尚未造成广泛影响的疑似虚假信息,平台在接到通知核实期间,可先行采取“打上存疑标签”“限制算法推荐流量(降权)”等柔性措施;对于已被证伪的谣言,除了删除外,平台还应采取“向已阅读该谣言的用户定向推送辟谣信息”的措施,以消除不良影响。此外,平台还应从技术底层将生成该谣言的特定“恶意提示词”加入黑名单,防止同类谣言被反复生成。

其次,建立针对重大谣言的快速响应与绿色通道机制。对于涉及自然灾害、公共卫生事件、金融安全等极易引发社会恐慌的生成式人工智能谣言,常规的“通知-审核-处置”流程耗时过长。可以要求平台建立针对此类重大谣言的快速响应机制,在接到相关部门或权威机构的通知后,必须在极短的时间内(如数小时内)采取阻断措施,有效切断谣言的传播链条,将社会危害降至最低。

基金项目

重庆工商大学研究生科研创新项目“生成式人工智能网络谣言法律责任研究”(yjscxx2025-269-137)。

参考文献

- [1] 翟月荧. 网络谣言的传播与治理[J]. 东岳论丛, 2023, 44(8): 150-156.
- [2] 李政葳. 我国生成式人工智能用户规模达 6.02 亿人[N]. 光明日报, 2026-02-06(004).
- [3] 魏思佳. “有图有真相”的涉灾谣言如何破——浅析 AI 技术带来的挑战和机遇[J]. 中国应急管理, 2024(8): 92-95.
- [4] 付阳. AI 时代网络谣言传播问题及对策分析[J]. 新闻研究导刊, 2024, 15(19): 10-13.
- [5] 梁市. AI 谣言传播特点、治理困境及对策[J]. 新媒体研究, 2024, 10(23): 6-12.
- [6] 黄河. 网络谣言的智能化演变及治理[J]. 人民论坛, 2023(4): 62-65.
- [7] 向安玲, 张诗瑶, 程栋. AI 谣言传播之组态路径研究——基于 35 个案例的定性比较分析[J]. 未来传播, 2025, 32(1): 46-56.
- [8] 黄玥, 黄泽森. 从 Sora 看生成式 AI 虚假信息引发的网络舆情风险及应对治理策略[J]. 广东公安科技, 2024, 32(4): 12-14.
- [9] 陆建平, 等. AIGC 时代虚假信息的制造传播与国家安全及公民权益的维护[J]. 浙江大学学报(人文社会科学版), 2024, 54(5): 42-58.
- [10] 张永亮. 人工智能的场景化应用及其法律规制研究[M]. 杭州: 浙江大学出版社, 2024.
- [11] 吴汉东. 人工智能时代的制度安排与法律规制[J]. 法律科学(西北政法大学学报), 2017, 35(5): 128-136.
- [12] 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.
- [13] 梁远高. 生成式人工智能服务提供者侵权责任的场景化分类及归责认定[J]. 深圳大学学报(人文社会科学版), 2024, 41(5): 115-124.
- [14] 陈亮, 张翔. 欧盟生成式人工智能立法实践及镜鉴[J]. 法治研究, 2024(6): 105-118.
- [15] 漆晨航. 生成式人工智能的虚假信息风险特征及其治理路径[J]. 情报理论与实践, 2024, 47(3): 112-120.
- [16] 朱振. 归责何以可能: 人工智能时代的自由意志与法律责任[J]. 比较法研究, 2022(1): 39-54.
- [17] 程啸. 侵权责任法[M]. 第 3 版. 北京: 法律出版社, 2021.
- [18] 沈森宏. 论生成式人工智能服务提供者过错的认定[J]. 现代法学, 2024, 46(6): 133-147.
- [19] 郭金良. 生成式人工智能服务中“信息错误”的民事责任[J]. 政法论坛, 2025, 43(2): 25-35.