ISSN 2095－9206
CN 10－1281/D

财经法学
Caijing Faxue

2026 年 3 月 15 日出版　　2026 年第 2 期（总第 68 期）　双月刊

□ **财经法治热点：**数字法治研究

论生成式人工智能侵害名誉权的侵权责任 …… 程　啸　3
网络时代稳定币的货币地位与规范研究 …… 刘少军　22
论作为独立财产保护的人工智能模型参数
——以“AI 模型保护第一案”为切入点 …… 廖慧姣　36
公共数据资源登记的制度缺漏及因应 …… 邓　鹏　53

□ **专论**

论投资风险对赌条款效力的二元论裁判思维：兼议公司法司法解释的
体系化 …… 刘俊海　69
多数人侵权责任的规范困境与出路 …… 张平华　90
走向台前的相对所有权
——兼及所有权和所有制的关系 …… 曹相见　107
“内卷式”竞争治理中反垄断法的制度性回避与功能重构 …… 喻　玲　125
商事逻辑下合同诈骗罪的认定范式 …… 毛玲玲　141
《个人所得税法》中偶然所得条款的适用反思与要件重构 …… 李乔彧　159

□ **争鸣**

金融借贷利率上限司法规制的路径重构 …… 钱　进　176
帮助信息网络犯罪活动罪“主观明知”认定规则检视 …… 谢甜甜　193

期刊基本参数：CN 10－1281/D*2015*b*16*208*zh*P*￥50.00*350*12*2026－03

LAW AND ECONOMY

No. 2, 2026 (Serial No. 68)

March 15, 2026

On Generative Artificial Intelligence's Tort Liability for Infringement of Right to Reputation ····· *Cheng Xiao* (3)

The Studies on Monetary Status and Regulation of Stablecoins in the Network Era ····· *Liu Shaojun* (22)

AI Model Parameters as Independent Property: Taking First Case on AI Model Protection as a Starting Point ····· *Liao Huijiao* (36)

Institutional Gaps and Responses in the Registration of Public Data Resources ····· *Deng Peng* (53)

Binary Judicial Philosophy on the Validity of Investment Risk Gamble Clauses: Also Discussing the Systematization of Judicial Interpretations of Company Law ····· *Liu Junhai* (69)

Normative Predicaments and Core Solutions of Joint and Several Liability for Torts ····· *Zhang Pinghua* (90)

The Emerging Relative Ownership: Also on the Relationship between Ownership and Ownership System ····· *Cao Xiangjian* (107)

Institutional Avoidance and Functional Reconstruction of Antitrust Law in Governing "Involutionary" Competition ····· *Yu Ling* (125)

The Paradigm of Contract Fraud Crime under Commercial Logic ····· *Mao Lingling* (141)

Revisiting the Application and Reconstructing the Elements of "Incidental Income" in China's Individual Income Tax Law ····· *Li Qiaoyu* (159)

The Reconstruction of Regulatory Control of the Upper Limit of Financial Lending Rates ····· *Qian Jin* (176)

Examination of the Identification Rules for "Subjective Knowledge" in the Crime of Aiding Information Network Criminal Activities ····· *Xie Tiantian* (193)

财经法学 No. 2，2026
pp. 3 - 21

论生成式人工智能侵害名誉权的侵权责任

程　啸*

内容提要：生成式人工智能服务提供者对于保证其输出内容的真实性和准确性具有技术上的控制力与法律义务。生成式人工智能的内容输出是提供者自身所实施的行为。提供者并非单纯的网络技术服务提供者，不能援引网络侵权责任中的通知规则与知道规则免除责任，而应依据过错责任原则为自己的行为负责任。生成式人工智能直接针对其名誉被涉及的用户的内容输出不构成发布行为，其输出的内容只有为被侵权人之外的第三人知悉时，该内容的输出才构成发布行为。该类输出内容既包括事实陈述，也包括意见表达，提供者可以主张真实性抗辩与公正评论抗辩。提供者对于侵害名誉权的过错包括故意和过失。对于过失的判断，应当首先依据法律规定的义务适用违法视为过失规则；没有法律规定的义务或者符合法定要求时，应当结合案件具体事实，根据《民法典》第998条的动态系统论与第1026条列举的考虑因素，按照侵害行为发生时的技术发展水平来认定提供者是否具有过失。

关键词：生成式人工智能　提供者　名誉权　侵权责任　过错

一、问题的提出

生成式人工智能（generative artificial intelligence）是一类能够根据输入数据自动生成新内容的人工智能系统，它通过学习大量现有数据中的模式、结构和特征，能够创造出新的文本、图像、音频、视频等形式的内容。〔1〕在目前世界上具有代表性的生成式人工智能中，文本生成类主

* 程啸，清华大学法学院教授。
本文为国家社科基金重点项目“人工智能侵权责任研究”（25AFX006）的阶段性成果。
感谢王苑、李西泠、杨嘉祺对本文初稿提出的修改意见以及林琳同学协助整理相关文献资料。

〔1〕《生成式人工智能服务管理暂行办法》第2条第1款将生成式人工智能服务界定为“利用生成式人工智能技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等内容的服务”。

要包括 OpenAI 的 ChatGPT、Anthropic 的 Claude、Google 的 Gemini、Meta 的 LLaMA、深度求索的 DeepSeek、阿里的千问、字节跳动的豆包及百度的文心一言；图像视频生成类包括 OpenAI 的 DALL-E、Sora，Google 的 Imagen、Veo，腾讯混元，阿里的通义万相等。生成式人工智能的核心功能在于创作与生成。因此，在其输出的文本、视频或图像等内容存在虚假错误时，可能产生侵害名誉权（及其他人格权）侵权责任。具体而言，可以分为两类情形：一是，生成式人工智能输出的内容是虚假错误的而直接侵害他人的名誉权，由此产生人工智能公司承担侵权责任的问题；二是，用户利用人工智能生成具有诽谤性或侮辱性的内容而侵害他人名誉权，此时，涉及用户与人工智能公司应否及如何承担侵权责任的问题。例如，在 2023 年美国发生的“Mark Walters v. OpenAI，L. L. C. 案”中，原告是公众人物、全国联合脱口秀主持人马克·沃尔特斯（Mark Walters），其以 ChatGPT 生成不实内容（谎称其被指控挪用第二修正案基金会的资金）为由向法院起诉运营 ChatGPT 的 OpenAI 公司构成诽谤侵权。〔2〕美国联邦地区法院佐治亚州北区亚特兰大分院经审理后以 ChatGPT 关于原告挪用公款的陈述不能被合理解读为事实陈述，且未能证明 OpenAI 公司存在过错等理由驳回了原告的起诉。〔3〕再如，我国北京互联网法院审理的一个案件中，被告孙某利用生成式人工智能将原告程某用作微信头像的肖像照创作生成为乳房暴露的动漫风格图片，并将该图片发送至与原告同在的摄影交流微信群内。在原告多次制止后，被告仍继续使用人工智能软件将原告的上述微信头像的肖像照片生成乳房暴露且身体畸形的动漫风格图片，并一对一发送给原告。法院判决认为“涉案 AI 软件是一种 AI 工具，被告是该工具的使用者，被诉侵权图片是被告自主决定生成，并在涉案微信群中发布。被告明知在人数众多的微信群中发布被诉侵权图片，会导致原告社会评价降低的后果，经原告劝阻，仍有意为之，其侵权故意明显”，故此，认定被告生成并发送涉案图片到微信群的行为构成对原告名誉权的侵害。〔4〕

目前，理论界对生成式人工智能侵权责任的归责原则、损害、过错、因果关系等问题已展开了一定的研究，〔5〕但尚未见到对生成式人工智能侵害名誉权的专门讨论。有鉴于此，本文拟对生成式人工智能侵害名誉权的侵权责任加以分析研究，主要研究以下问题：首先，生成式人工智能的内容输出能否被认定为生成式人工智能服务提供者的行为，抑或提供者作为网络技术服务提供者而只是单纯的传输他人生产或形成的内容或信息；〔6〕其次，在确认生成式人工智能的内容输出是提供者的行为之后，进一步分析如何认定输出的内容构成侵害名誉权的加害行为与产生了名誉权被侵害的后果；最后，分析在生成式人工智能侵害名誉权的案件中怎样判断提供者有无过错的问题。〔7〕

---

〔2〕 See Mark Walters *v*. OpenAI，L. L. C.，1：23-cv-03122，(N. D. Ga.).

〔3〕 该案详细资料参见 https://www. courtlistener. com/docket/67617826/walters-v-openai-llc/，2025 年 11 月 20 日访问。

〔4〕 参见北京互联网法院（2024）京 0491 民初 21163 号民事判决书。

〔5〕 代表性研究如王利明：《生成式人工智能侵权的归责原则与过错认定》，载《中国法律评论》2025 年第 4 期；王利明：《生成式人工智能侵权的法律应对》，载《中国应用法学》2023 年第 5 期；王若冰：《论生成式人工智能侵权中服务提供者过错的认定——以“现有技术水平”为标准》，载《比较法研究》2023 年第 5 期；徐伟：《论生成式人工智能服务提供者侵权损害赔偿》，载《财经法学》2025 年第 2 期；周学峰：《生成式人工智能侵权责任探析》，载《比较法研究》2023 年第 4 期；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》，载《政法论坛》2025 年第 6 期。

〔6〕 生成式人工智能服务提供者，是指利用生成式人工智能技术提供生成式人工智能服务（包括通过提供可编程接口等方式提供生成式人工智能服务）的组织、个人（《生成式人工智能服务管理暂行办法》第 22 条第 2 项）。提供者可能就是人工智能系统的研发者，也可能是研发者之外的运营者。

〔7〕 因篇幅所限，对于因果关系、损害等问题，笔者将另行撰文研究。

二、内容输出是否是生成式人工智能服务提供者的行为

生成式人工智能的内容输出是否属于提供者实施的行为这一问题，可以细分为两个层次的小问题：首先，该内容之输出究竟是归属于生成式人工智能系统本身还是提供者；其次，如果归属于提供者，该内容的输出只是单纯的传输这一技术服务行为，还是应当被认定为提供者自身的言论发布行为。前一问题涉及行为主体的确定，后一问题则是对行为性质的认定。理论界曾有将人工智能本身作为一个法律主体看待并赋予法律人格的观点。〔8〕例如，2017 年欧洲议会通过的《欧洲机器人技术民事法律规则》第 59 条（f）认为："从长远来看要创设机器人的特殊法律地位，以确保至少最复杂的自动化机器人可以被确认为享有电子人的法律地位，有责任弥补自己所造成的任何损害，并且可能在机器人作出自主决策或以其他方式与第三人独立交往的案件中适用电子人格。"基于这种观点，生成式人工智能所输出的内容不是人工智能创造者（即人工智能公司）的言论，甚至根本就不是人类言论，而是人工智能本身的言论。〔9〕然而，无论是赋予人工智能以伦理性人格还是技术性人格，目前都是不可行的。〔10〕一方面，如果将自我意识作为"自主性"不可或缺的要素，则目前的通用型人工智能虽然具有相当的自主性，但距离自然人那样的可以广泛地理解、学习和执行各种任务的认识思维能力水平还有很远的距离，遑论具有人类心灵的能力。〔11〕"只有当机器能够模拟主客观世界的不确定性，使定性的人类思维可以用带有不确定性的定量方法去研究，才能最终使机器具有更高的智能，在不同尺度上模拟和代替人脑的思维活动。"〔12〕因此，无法将人工智能作为自然人那样看待，赋予其伦理人格。另一方面，人工智能没有自己独立的财产，无法承担法律责任，更没有自然人那样的意思表示机制，因此，法律上无法赋予人工智能以法人那样的拟制人格（即技术人格），最终还是需要明确究竟谁是行为人与责任人。总之，只要人工智能不具有完全的自主性，仍是由人类来创建人工智能或至少是决定使用人工智能，只要人工智能自身还不拥有独立的财产，无法独立承担责任，则行为与责任总是要归属于现行法律所承认的自然人、法人或非法人组织等民事主体。生成式人工智能也不例外。因此，对于上述第一个问题回答是：生成式人工智能输出内容应当归属于提供者，是提供者的行为。

上述第二个问题，即输出内容的行为究竟是单纯的传输这一技术性服务还是提供者自身发表言论的行为，对于提供者如何承担侵害名誉权的侵权责任至关重要。如果输出的内容就是提供者发表的言论，则其应当按照过错责任承担侵权责任。倘若只是单纯的传输行为，那么，提供者可以援引网络侵权责任中的通知规则与知道规则而免责，只在未及时采取必要措施的情况下才需要承担责任。对于该问题，目前国内外法学界存在很大的争议。下面先介绍比较法和我国法上的争

〔8〕 相关研究比如袁曾：《人工智能有限法律人格审视》，载《东方法学》2017 年第 5 期；刘云：《论人工智能的法律人格制度需求与多层应对》，载《东方法学》2021 年第 1 期；石冠彬：《人工智能成为法律主体不存在理论障碍》，载《光明日报》2024 年 8 月 16 日，第 11 版。

〔9〕 See Peter N. Salib，*AI Ouputs Are Not Protected Speech*，102 Washington University Law Review 83（2024）.

〔10〕 参见梅夏英：《伦理人格与技术人格：人工智能法律主体地位的理论框架》，载《中外法学》2025 年第 1 期，第 42－43 页。

〔11〕 参见冯珏：《自动驾驶汽车致损的民事侵权责任》，载《中国法学》2018 年第 6 期，第 112 页。

〔12〕 李德毅等：《不确定性人工智能》，载《软件学报》2004 年第 11 期，第 1590 页。

论，然后阐述本文的观点。

（一）比较法上的争议

1. 美国法

1996 年的美国《通信规范法》第 230 条（c）款为“交互式计算机服务的提供者或使用者”提供了豁免权。该款明确规定，交互式计算机服务的提供者或使用者不得被视为另一信息内容提供者所提供信息的发布者或传播者，因此，无须因为以下行为承担责任：（1）出于善意自愿采取行动，限制对其认为属于淫秽、猥亵、淫荡、污秽、极端暴力、骚扰性或其他令人反感材料的访问或获取（无论该材料是否受宪法保护）；（2）采取行动，为信息内容提供者或他人提供技术手段，以限制对第（1）项所述材料的访问。上述规定使得平台免于为第三方内容承担责任，被认为是现代互联网发展的重要法律基础。〔13〕司法实践中，美国法院逐渐明确了《通信规范法》第 230 条的适用须满足以下三个条件：第一，被告必须是交互式计算机服务的提供者或使用者；第二，索赔必须基于另一信息内容提供者提供的信息；第三，索赔必须将被告视为发布者。〔14〕基于对这三个限制条件的理解，美国法学界就生成式人工智能服务的提供者（也称运营者）能否适用《通信规范法》第 230 条的问题存在肯定说、否定说与区分适用说等三种不同观点。

肯定说认为，生成式人工智能服务符合《通信规范法》第 230 条的定义，可以适用该条。首先，法院普遍认为，允许用户访问托管于服务器的内容或服务的网站即属于“交互式计算机服务”，生成式人工智能平台允许用户访问其托管的人工智能模型，符合该定义。其次，生成式人工智能是基于从互联网、授权方和用户处获取的第三方海量数据进行训练后，根据用户提示，通过识别和重组这些信息中的模式来生成回应。生成式人工智能获取这些信息后不会存储所依赖的信息，相关信息始终归第三方所有。生成式人工智能是通过梳理第三方内容，响应用户提示生成内容。除了用户的输入外，生成式人工智能并未自主创造新的内容。再次，与图书馆或报摊类似，生成式人工智能响应用户输入整合第三方内容时，扮演的正是“发布者”的角色。因此，针对人工智能公司的诉讼大概率会将其视为信息发布者，如主张人工智能生成诽谤内容的诉讼就会要求法院将人工智能视为该信息的发布者。〔15〕最后，尽管 ChatGPT 等生成式人工智能技术非常复杂，但技术的复杂性不改变生成式人工智能作为“中性工具”的法律本质。法院在判断《通信规范法》第 230 条的可适用性时关注的是平台是否在扮演发布者角色而非其算法的复杂程度。并且，国会立法意图也支持对第 230 条豁免作宽泛解释，从而保护技术平台免于为无法控制的第三方内容承担过重的责任。〔16〕

否定说认为，生成式人工智能不能适用《通信规范法》第 230 条。首先，该条要求被告是交

〔13〕 See Eric Goldman, *Why Section 230 Is Better Than the First Amendment*, 95 Notre Dame Law Review Reflection 33 (2019).

〔14〕 See Kathleen Ann Ruane, *How Broad a Shield? A Brief Overview of Section 230 of the Communications Decency Act*, Congressional Research Service (February 2018), https://www.congress.gov/crs_external_products/LSB/PDF/LSB10082/LSB10082.2.pdf, visited on 18 November 2025.

〔15〕 See Louis Shaheen, *Section 230's Immunity for Generative Artificial Intelligence*, 15 Seattle Journal of Technology, Environmental, & Innovation Law 1, 17 (2024).

〔16〕 See Louis Shaheen, *Section 230's Immunity for Generative Artificial Intelligence*, 15 Seattle Journal of Technology, Environmental, & Innovation Law 1, 16-20 (2024).

互式计算机服务的提供者或使用者，而非违法内容的提供者。人工智能程序的输出是由程序自身生成的，并非简单地引用现有网站内容（如搜索引擎提供的网站摘要）或现有用户查询内容（如某些自动补全功能通过引用用户提供的内容推荐下一个或多个词汇）。因此，针对人工智能公司的诉讼，本质上是要求将人工智能公司视为信息的发布者或传播者，该公司就是潜在的责任主体——“信息内容提供者”。其次，法院此前明确了第230条不豁免那些对有害内容有“实质性贡献”的被告，而人工智能公司创造并提供了生成不当内容的程序，其对生成内容具有实质贡献。再次，生成式人工智能看似基于训练数据中已有的词汇生成内容，但就像某人通过复制粘贴他人发布的每一个字（而非亲自写每一个字）来构成诽谤性文本时被认为同样创作了相关内容一样，人工智能生成的诽谤性文本不能被认为只是对已有内容的简单聚集和复制粘贴，而是对其诽谤性质产生了实质贡献。最后，适用《通信规范法》第230条的传统案例至少都在理论上让言论的实际创造者为其负责。如果让人工智能程序的生成内容可以获得第230条的豁免，则一般性地完全切断了受害人向任何人寻求救济的可能。〔17〕

区分适用说认为，就能否适用《通信规范法》第230条的问题，不能一概而论，要区分不同情况。在区分适用说中，有的学者认为，关键的问题在于平台究竟只是发布第三方的内容还是其自身就是“信息内容提供者”，即生成式人工智能是否对作为原告侵权赔偿诉讼基础的可诉内容进行了“创作”或“开发”。判断该问题时，法院主要关注的是平台与内容的互动性质。如果平台对违法内容的产生有实质促进，则不享有第230条的豁免权。〔18〕据此，倘若生成式人工智能存在以下情形，则法院可能拒绝给予其第230条豁免权：对内容的违法性因素负有责任从而对内容构成实质性促成；超出“传统编辑功能”的范围而添加了新内容，不再是第三方信息的“中介”；属于违法内容的“作者”，而非仅展示第三方违法内容；索赔依据聚焦于被告的其他行为或商业惯例，而非平台上展示的内容。如果法院认定存在以下情形，则生成式人工智能可能享有第230条的保护：仅仅转发、整理或总结现有第三方内容且对内容的修改属于“传统编辑功能”，基础信息完全是由第三方提供的且通过“中性”的算法进行过滤或处理。〔19〕还有的学者认为，应当结合人工智能的技术设计细节进行具体分析：其一，模仿预训练数据中的语言模式生成新文本的基准基础模型（baseline foundation models）会响应提示生成新内容，而非仅托管或传递他

〔17〕 See Eugene Volokh, *Large Libel Models? Liability for AI Output*, 3 Journal of Free Speech Law 489, 494 - 498 (2023). 2023年6月，美国参议员提出议案，主张将生成式人工智能排除在《通信规范法》第230条规定的豁免权之外。See *S. 1993-A Bill to Waive Immunity under Section 230 of the Communications Act of 1934 for Claims and Charges related to Generative Artificial Intelligence*, https://www.congress.gov/bill/118th-congress/senate-bill/1993, visited on 8 November 2025; Julia Shapero, *Hawley Calls for Repeal of Tech Legal Shield as AI Rises*, (9 May 2025), https://thehill.com/policy/technology/5488200-repeal-section-230-law-tech/, visited on 8 November 2025.

〔18〕 See Jason Davidson & Hilary G. Buttrick, *SAY WHAT?! When ChatGPT Gets it Wrong: Examining Generative AI, Section 230 of the Communications Decency Act, and the Essence of Creativity*, 30 Richmond Journal of Law and Technology 143, 167 (2024).

〔19〕 See Jason Davidson & Hilary G. Buttrick, *SAY WHAT?! When ChatGPT Gets it Wrong: Examining Generative AI, Section 230 of the Communications Decency Act, and the Essence of Creativity*, 30 Richmond Journal of Law and Technology 143, 178 (2024). 类似观点认为，在某些产品整合场景中，若生成式人工智能仅用于确定搜索结果的优先级，或从基础搜索结果中高亮显示特定文本，法院可能会认定其有权享受第230条的保护。而人们讨论的许多大语言模型使用场景，则超出了第230条的保护范围。See Matt Perault, *Section 230 Won't Protect ChatGPT*, 3 Journal of Free Speech Law 363, 366 (2023).

人创作信息，因此不能获得第 230 条的豁免；其二，输出的内容直接取自第三方来源的抽取式生成（extractive generation）的情况下，因其与搜索引擎摘要等已获豁免的功能高度相似，豁免的可能性较高；其三，检索增强（retrieval-augmented）生成模型虽然借鉴了搜索功能的特性，但可能生成未基于检索源的虚构内容，其能否豁免存在不确定性；其四，基于人类反馈进行学习（learning from human feedback）的模型更接近平台自主创作内容，豁免的概率会降低。〔20〕

从美国的司法实践来看，在联邦第二巡回法院审理的 Force v. Facebook 案中，法院审查的核心问题是《通信规范法》第 230 条是否阻却作为美国公民的受害者及其近亲属就 Facebook 使用一种基于用户输入来决定向其展示何种内容的算法而向该平台提起诉讼的可能性。法院认为，使用算法本身并不会使交互式计算机服务成为发布者，因为 Facebook 的算法只是组织第三方信息。所谓信息内容提供者，是指对通过互联网或任何其他交互式计算机服务提供的信息的创建或开发全部或部分负责的任何个人或实体。如果 Facebook 创建或开发了原告索赔所依赖的内容，那么 Facebook 就可能被追究责任，因为这将把 Facebook 的身份从“发布者”转变为“信息内容提供者”，并剥夺其受第 230 条保护的权利。在联邦第二巡回法院看来，Facebook 的算法只是获取用户提供的信息，并将该内容与其他第三方内容匹配从而生成供用户消费的信息流。Facebook 只是显示未经更改的内容并使其更可见而已，故而，不应被视为第 230 条规范目的下的内容创建者。〔21〕

2. 欧盟法

2000 年欧盟颁布的《电子商务指令》（Directive 2000/31/EC）的第四节对“中间服务提供者的责任”作出了规定。该法第 12—14 条分别规定了在满足特定条件时，服务提供者不因提供纯粹传输服务，对信息的自动、中间性和暂时的存储以及根据用户的要求存储信息而承担责任。此外，第 15 条还进一步明确了成员国不应当要求服务提供者承担监督其传输和存储的信息的一般性义务，也不应当要求服务提供者承担主动收集表明违法活动的事实或情况的一般性义务，但是，可以要求服务提供者承担立即向主管公共机构报告其服务接受者进行的非法行为或者提供的非法信息的义务，或者应主管当局的要求而提供可以确定与其有存储协议的服务接受者的身份信息的义务。〔22〕2022 年欧盟又颁布了《数字服务法》（Digital Services Act，DSA）。该法继续保留了 2000 年《电子商务指令》有关中间服务提供者责任有条件豁免的法律框架，并进一步丰富完善了相关规则以适应数字经济的发展。与美国一样，欧盟学者对于生成式人工智能服务提供者能否适用《数字服务法》关于中间服务提供者责任豁免的规定也存在争议，有肯定说、部分肯定说与否定说等不同观点。

肯定说认为，人工智能系统自主生成的内容应当视为人工智能系统运营者的行为。首先，运营者通过选择应用领域决定了内容的传播语境，进而影响训练数据的流入。其次，这些系统越来

〔20〕 See Peter Henderson, Tatsunori Hashimoto & Mark Lemley, *Where's the Liability in Harmful AI Speech?*, 3 Journal of Free Speech Law 589, 622 - 626 (2023).

〔21〕 See Louis Shaheen, *Section 230's Immunity for Generative Artificial Intelligence*, 15 Seattle Journal of Technology, Environmental, & Innovation Law 1, 14 - 16 (2024).

〔22〕 See *Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market* (*'Directive on Electronic Commerce'*), https://eur-lex.europa.eu/eli/dir/2000/31/oj/eng, visited on 8 November 2025.

越多地被运营者用作辅助工具，如客户沟通，这也表明这些内容可视为运营者的自有内容。因此，至少原则上可以将责任归属于运营者，按照一般原则对自有内容负责。〔23〕再次，在2013年5月14日德国联邦最高法院（BGH）第六民事审判庭针对谷歌自动补全功能的判例中，法院原则上认定谷歌对非法搜索词补全负有责任，但通常仅在运营者违反反应性审查义务时才承担责任。根据这一判例，原则上可以认定通信人工智能能够像人类通信一样侵害人格权。因此，在通信人工智能的言论可以归责于服务提供商的情况下，属于《电信媒体法》第7条第1款意义上的其“自己的信息”，提供者需要根据一般法律负责。〔24〕

部分肯定说认为，存在一些可行的解释方式能够将特定使用情境的生成式人工智能纳入《数字服务法》的适用范围，此说主要分为搜索引擎路径说和托管路径说。搜索引擎路径说尝试通过将生成式人工智能解释为搜索引擎，进而使之落入《数字服务法》的监管范围，理由在于：首先，搜索引擎难以直接划为纯粹传输、缓存或托管这几类“中间服务”中的某一种，但仍然受《数字服务法》规制。这说明《数字服务法》对中间服务各类别的定义是灵活的。〔25〕其次，搜索引擎的核心功能在于筛选并排序“极有可能满足用户需求的在线资源列表”，有些生成式人工智能被用于在信息富集的环境中组织信息，两者具有功能上的高度相似性。〔26〕再次，有些生成式人工智能被嵌入已受《数字服务法》监管的现有平台，此时，该类生成式人工智能视具体情况有可能被作为搜索引擎的一部分而适用《数字服务法》，如嵌入必应搜索引擎中的必应（Bing）聊天采用GPT-4生成答案及创意内容，可被视为必应搜索引擎的有机组成部分，进而可能需作为必应搜索引擎的一部分遵守《数字服务法》的所有义务。〔27〕托管路径说则认为，生成式人工智能可以解释为托管服务（hosting），理由在于：首先，生成式人工智能的运作包含“生成输出”之外的多个阶段。在判断人工智能生成内容的“实际创作者”时，应当参考人类与人工智能在内容制作中的参与程度。这可能因个案差异有所不同，需要单独评估每个具体的使用场景，无法直接将所有生成式人工智能应用一概归类为“实际内容创作者”。〔28〕其次，生成式人工智能可能符

〔23〕 Vgl. Hotz, Thorsten, Persönlichkeitsrechtliche Haftung beim Einsatz autonomer Systeme, ZUM 2025, S. 89.

〔24〕 Vgl. Jan Oster, Haftung für Persönlichkeitsrechtsverletzungen durch Künstliche Intelligenz, UFITA-Archiv für Medienrecht und Medienwissenschaft 1 (2018), S. 14 - 52.

〔25〕 See Sophie Stalla-Bourdillon, *What if ChatGPT Was Much More than a Chatbox? What if LLM-as-a-service Was a Search Engine?*, https://peepbeep.blog/2023/04/03/what-if-chatgpt-was-much-more-than-a-chatbox-what-if-llm-as-a-service-was-a-search-engine/, visited on 8 November 2025.

〔26〕 See Beatriz Botero Arcila, *Is It a Platform? Is It a Search Engine? It's ChatGPT! The European Liability Regime for Large Language Models*, 3 Journal of Free Speech Law 455 (2023). 相似观点认为，部分大型语言模型基于固定数据集运行，并未“对所有网站进行搜索”，因此不符合在线搜索引擎的定义。但ChatGPT等部分大型语言模型会整合当前网页的搜索结果，可能符合搜索引擎的定义。See Mathias Vermeulen & Laureline Lemoine, *From ChatGPT to Google's Gemini: When Would Generative AI Products Fall within the Scope of the Digital Services Act?*, (12 February 2024), https://blogs.lse.ac.uk/medialse/2024/02/12/from-chatgpt-to-googles-gemini-when-would-generative-ai-products-fall-within-the-scope-of-the-digital-services-act/, visited on 5 November 2025.

〔27〕 See Mathias Vermeulen & Laureline Lemoine, *From ChatGPT to Google's Gemini: When would Generative AI Products Fall within the Scope of the Digital Services Act?*, (12 February 2024), https://blogs.lse.ac.uk/medialse/2024/02/12/from-chatgpt-to-googles-gemini-when-would-generative-ai-products-fall-within-the-scope-of-the-digital-services-act/, visited on 8 November 2025.

〔28〕 See Ioannis Revolidis, *Generative AI Content Misuse and the DSA*, *In XIX IDP Conference (Internet, Law, and Politics): Current Challenges of Artificial Intelligence*, Open University of Catalonia, 2024, pp. 83 - 96.

合欧洲法院在“谷歌搜索案”（Google Search）及近期判例中采用的“广义托管服务定义”，〔29〕它们通常会在服务器中存储“查询与用户信息”，并且内容生成过程具有纯技术性、自动化特征且对所存储或生成的内容缺乏认知与控制，具有中立性。〔30〕最后，生成式人工智能会在服务器内存中临时存储用户提示词（输入内容）等特定信息。模型输入可被视为用户提供，并应用户要求由大型语言模型存储。由于这些输出是由用户的提示词所触发的，输出结果也可被认定为用户“提供”的行为。〔31〕

否定说认为，生成式人工智能不适用《数字服务法》。这是因为，生成式人工智能通常不属于《数字服务法》所规定的“中间服务”中的“纯粹传输”“缓存”或“托管”中的任何一类。最接近的可能是托管服务，但针对这类情况，根据欧洲法院的观点，即使是仅存储用户生成内容的平台，若偏离“中立地位”，也可能失去其在《数字服务法》下的“托管服务提供者”地位。例如，平台若对特定用户生成内容进行推广，就可能被认定为偏离中立地位。在生成式人工智能场景，正是人工智能模型本身提供了内容，因此对于自行生成内容的系统开发者，更不能将其视为托管服务提供者。〔32〕此外，《数字服务法》制定的初衷也并不旨在应对生成式人工智能产生的内容。〔33〕

（二）我国学界的不同观点

依据我国《民法典》第 1195—1197 条确立的包括通知规则与知道规则在内的网络侵权规则，当网络服务提供者并未直接实施侵害他人民事权益的行为，而只是网络用户利用网络服务提供者所提供的网络服务实施侵权行为时，网络服务提供者原则上不需要为他人的侵权行为承担责任。只有在以下两种情形中，网络服务提供者才需要承担侵权责任：其一，网络服务提供者知道或应当知道网络用户利用网络服务实施侵权行为而未及时采取必要措施的，应当与该网络用户承担连带责任（《民法典》第 1197 条），这就是所谓的“知道规则”；其二，网络服务提供者不知道且不应当知道网络用户利用生成式人工智能服务实施侵权行为，但是，其在收到权利人的通知后未及时采取必要措施的，需要就损害的扩大部分与该网络用户承担连带责任（《民法典》第 1195—1196 条）这就是所谓的“通知规则”，也称“通知删除规则”。

通知规则与知道规则是《民法典》为了协调权益保护与行为自由，避免过度加重网络服务提供者的审查义务、妨害信息网络科技发展，而在借鉴美国等国外相关立法的基础上确立的适用于网络技术服务提供者的独特侵权规则。〔34〕除非本地部署且仅通过局域网供内部人员使用，否则，

---

〔29〕 See Google France SARL *v*. Louis Vuitton Malletier SA, [2010] ECR I-2417, (Mar. 23, 2010).

〔30〕 See Beatriz Botero Arcila, *Is It a Platform? Is It a Search Engine? It's ChatGPT! The European Liability Regime for Large Language Models*, 3 Journal of Free Speech Law 455 (2023).

〔31〕 See Mathias Vermeulen & Laureline Lemoine, *From ChatGPT to Google's Gemini: When would Generative AI Products Fall within the Scope of the Digital Services Act?*, (12 February 2024), https://blogs.lse.ac.uk/medialse/2024/02/12/from-chatgpt-to-googles-gemini-when-would-generative-ai-products-fall-within-the-scope-of-the-digital-services-act/, visited on 2 November 2025.

〔32〕 Vgl. Berz, Amelie/Engel, Andreas/Hacker, Philipp: Generative KI, Datenschutz, Hassrede und Desinformation-Zur Regulierung von KI-Meinungen, ZUM 2023.

〔33〕 See Philipp Hacker, Andreas Engel & Marco Mauer, *Regulating Chat GPT and Other Large Generative AI Models*, in the 2023 ACM Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery, 2023, p. 15.

〔34〕 参见黄薇主编：《中华人民共和国民法典侵权责任编释义》，法律出版社 2020 年版，第 94 页；程啸：《侵权责任法》（第 3 版），法律出版社 2021 年版，第 499 页以下。

生成式人工智能服务都是通过互联网提供的并且无须附着于有体物之上，属于网络服务。因此，在生成式人工智能侵害名誉权的侵权纠纷中，生成式人工智能服务的提供者属于网络服务提供者。问题是，其究竟是网络技术服务提供者（内容的输出不能被看作是提供者的行为），还是网络内容服务提供者（该内容就是提供者的行为）。对此，我国学界存在不同的观点。

有的学者认为，生成式人工智能的提供者固然存在不同于传统的网络内容提供者之处，如相较于传统的内容提供者，通用人工智能提供者对于生成的内容控制力降低，并且通用人工智能的创作行为需要人类的提示词予以激发等，〔35〕但是，就 ChatGPT 等大语言模型的提供者而言，大语言模型自动生成的信息内容并非是其用户创作或发布的信息，而应当看作是人工智能系统提供者本身创作、发布的信息。〔36〕有的学者认为，因为生成式人工智能服务提供者对输出内容的控制力弱于网络内容服务的提供者，并且，在我国立法中鲜有关于网络内容服务提供者侵权责任的规则，所以，在实质意义上不宜将生成式人工智能服务提供者认定为网络内容服务的提供者，否则将对现有法律体系带来不必要的冲击。〔37〕还有的学者在分析生成式人工智能搜索服务提供者能否认定为网络技术服务提供者而适用避风港规则予以免责时指出，生成式人工智能搜索与传统搜索的最大区别在于其直接向用户提供搜索的回答，该回答中源自来源网站的部分也是经由人工智能模型生成的，即其对内容进行了编辑加工，而非单纯的提供网页链接。故此，不能将此种人工智能服务提供者认定为网络技术服务提供者。〔38〕

（三）内容输出是生成式人工智能服务提供者所实施的行为

《民法典》第 1195—1197 条确立的通知规则与知道规则适用的是网络技术服务提供者。网络技术服务提供者只是单纯地按照用户的指令在用户指定的两点或多点之间通过信息网络就该用户所提供或修改的内容自动提供网络接入、信息传输、自动存储、存储空间、信息搜索、链接服务。在这个过程中，信息或内容是由用户所创作或提供的，网络技术服务提供者本身并不参与该信息或内容的生成或对之进行编辑、加工、整理或更改。然而，生成式人工智能服务则完全不同，它并非单纯的信息传播链条的中立节点。虽然生成式人工智能需要依据用户的指令而输出内容，但其本身也是基于数据、算法和模型等复杂的技术而生成文本、图片、音频、视频、代码等新的信息或内容。在这个过程中，生成式人工智能服务提供者并非处于纯粹消极的、中立的第三方地位，而是实质地参与了内容或信息的生成，由此在法律上负有相应的义务以避免产生侵害他人合法权益的内容以及违法、有害内容。因此，本文认为，生成式人工智能服务的内容输出是提供者实施的行为，理由阐述如下：

1. 生成式人工智能服务提供者对于输出内容的真实性、准确性与合法性具有实质上的控制力。对此，可以从生成式人工智能必备的三大要素即数据、算法和模型的角度分别加以分析。一方面，生成式人工智能离不开海量的数据，数据决定了它能够生成何种内容（文本、图片、视

〔35〕 参见高阳：《通用人工智能提供者内容审查注意义务的证成》，载《东方法学》2024 年第 1 期，第 193 页。

〔36〕 参见周学峰：《生成式人工智能侵权责任探析》，载《比较法研究》2023 年第 4 期，第 130 页。

〔37〕 参见徐伟：《论生成式人工智能服务提供者的法律地位及其责任——以 ChatGPT 为例》，载《法律科学（西北政法大学学报）》2023 年第 4 期，第 74 页。

〔38〕 参见姚志伟：《生成式人工智能搜索服务提供者侵权免责制度》，载《环球法律评论》2025 年第 6 期，第 56－57 页。

频、代码等），也决定了生成内容的上限。生成式人工智能的开发者进行大语言模型训练的最主要的一类数据是对合法取得的数据标注后形成的数据集。生成式人工智能数据标注（generative artificial intelligence data annotation），是指通过人工操作或使用自动化技术机制，基于对提示信息的响应信息内容，将特定信息如标签、类别或属性添加到文本、图片、音频、视频或者其他数据样本的过程。〔39〕数据标注本身虽然并不改变数据的原始形态和内容，但是，却深刻地改变了数据的语义、内涵、用途以及价值。所谓“越人工，越智能”。数据标注是否科学、准确、合法、合伦理，从根本上决定了生成式人工智能输出的内容的风格、质量、准确与安全。当数据标注不准确、不规范、不合法或者掺杂了标注者的性别、种族、民族等歧视性、侮辱性因素时，经过此种标注数据训练出的模型就往往内生了侵害他人合法权益的风险。由此可见，从数据标注的环节，人工智能开发者或提供者实际上就已经在实质性地影响输出的内容。

另一方面，如果说数据是原料，算法就是过程，模型则是结果。生成式人工智能的研发者在特定数据集上运行一定的算法后得到的就是模型，生成式人工智能生成内容是基于深度学习技术的不同模型通过算法来处理数据后生成文本、图像、视频等不同内容的，如通过扩散模型、Transformer 模型生成文本、图像、视频，基于卷积神经网络模型进行图像识别，基于循环神经网络模型进行语言翻译等。在生成式人工智能中，算法和模型本身就直接参与了内容的创造。在 GPT、Stable Diffusion 这类模型，其核心算法如 Transformer、Diffusion Model 是通过学习海量数据中的联合概率分布，即数据中多个变量组成的向量的概率分布，对已有的数据进行总结归纳。〔40〕以 Transformer 模型生成文本的过程为例，其核心在于注意力机制，即通过计算编码器端的输出结果中的每个向量与解码器端的输出结果中每个向量的相关性，得出若干相关性分数，再进行归一化处理将其转化为相关的权重，用来表征输入序列与输出序列各元素之间的相关性。这一机制使得模型能够深度理解整个输入序列中每个词之间的关系和上下文含义，减少了对外部信息的依赖，更擅长捕捉数据和特征的内部相关性。〔41〕算法和模型对输出的内容是什么具有实质性影响。生成式人工智能的开发者通过算法已经决定了模型理解世界和生成内容的“概率规则”，而非在每次输出时逐一进行干预。输出内容本质上是开发者通过算法预设的“世界观”与用户即时提示共同作用的结果，其贡献是深嵌于模型本质之中的。因此，生成式人工智能的研发者需要通过设计和选择不同的算法架构与训练目标，约束模型参数的学习方向，以确保输出内容的真实性、准确性与安全性。

2. 正是因为生成式人工智能是研发者所创造的，研发者从数据、算法和模型三个维度上对输出内容都具有实质性的、技术上的控制力，所以，法律上才从人工智能的研发、部署、运营、管理等各个环节给研发者施加了相应的义务，要求其通过技术手段实现生成式人工智能的安全、可靠与可控，防止输出诽谤性、歧视性以及其他违法有害内容。

例如，欧盟《人工智能法案》对于通用人工智能模型的提供者义务作出了详细的规定，如：

〔39〕 参见国家市场监督管理总局、国家标准化管理委员会发布的《网络安全技术 生成式人工智能服务安全基本要求》（GB/T45654—2025）第 3.5 条；《网络安全技术 生成式人工智能数据标注安全规范》（GB/T45674—2025）第 3.3 条。

〔40〕 参见丁磊：《生成式人工智能：AIGC 的逻辑与应用》，中信出版集团 2023 年版，第 6 页。

〔41〕 同上注，第 77－78 页。

编制并不断更新该模型的技术文件，包括其培训和测试过程及其评估结果；编制、不断更新并向意图将通用人工智能模型纳入其人工智能系统的提供者提供信息和文件，该信息和文件应当使人工智能系统的提供者能够很好地了解通用人工智能模型的能力和局限性，并遵守本条例规定的义务等（第53条）。对于那些具有系统风险的通用人工智能模型的提供者，该法案还规定了更多的义务，如：应当根据反映先进技术水平的标准化协议和工具进行模型评估，包括对模型进行对抗测试并记录在案，以识别和降低系统性风险；评估和减轻欧盟层面可能存在的系统性风险，包括因开发、投放市场或使用具有系统性风险的通用人工智能模型而产生的系统性风险的来源；跟踪、记录并及时向人工智能办公室报告，并酌情向成员国主管机关报告严重事件的相关信息以及为解决这些问题可能采取的纠正措施；确保对具有系统性风险的通用人工智能模型和模型的物理基础设施提供足够水平的网络安全保护等（第55条）。

在我国，法律法规规章也对此作出了明确的规定。2025年新修订的《网络安全法》第20条第1款规定："国家支持人工智能基础理论研究和算法等关键技术研发，推进训练数据资源、算力等基础设施建设，完善人工智能伦理规范，加强风险监测评估和安全监管，促进人工智能应用和健康发展。"《生成式人工智能服务管理暂行办法》第4条明确要求：提供和使用生成式人工智能服务，应当遵守法律、行政法规，尊重社会公德和伦理道德，不得生成虚假有害信息等法律、行政法规禁止的内容；在算法设计、训练数据选择、模型生成和优化、提供服务等过程中应当采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视；尊重他人合法权益，不得危害他人身心健康，不得侵害他人名誉权等人格权益。同时，该办法还要求：提供者应当依法开展预训练、优化训练等训练数据处理活动，采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性（第7条）；在生成式人工智能技术研发过程中进行数据标注的，提供者应当制定符合《生成式人工智能服务管理暂行办法》要求的清晰、具体、可操作的标注规则（第8条）；开展数据标注质量评估，抽样核验标注内容的准确性，监督指导标注人员规范开展标注工作（第8条）；当提供者发现违法内容时，应当及时采取停止生成、停止传输、消除等处置措施，采取模型优化训练等措施进行整改，并向有关主管部门报告（第14条）。《互联网信息服务算法推荐管理规定》明确要求：算法推荐服务提供者应当落实算法安全主体责任，建立健全算法机制机理审核、科技伦理审查、用户注册、信息发布审核、数据安全和个人信息保护、反电信网络诈骗、安全评估监测、安全事件应急处置等管理制度和技术措施（第7条）；算法推荐服务提供者应当定期审核、评估、验证算法机制机理、模型、数据和应用结果等，不得设置诱导用户沉迷、过度消费等违反法律法规或者违背伦理道德的算法模型（第8条）。此外，《互联网信息服务深度合成管理规定》明确规定：深度合成服务提供者和使用者不得利用深度合成服务制作、复制、发布、传播虚假新闻信息，转载基于深度合成服务制作发布的新闻信息的，应当依法转载互联网新闻信息稿源单位发布的新闻信息（第6条第2款）；深度合成服务提供者应当落实信息安全主体责任，建立健全用户注册、算法机制机理审核、科技伦理审查、信息发布审核、数据安全、个人信息保护、反电信网络诈骗、应急处置等管理制度，具有安全可控的技术保障措施（第7条）；深度合成服务提供者应当加强深度合成内容管理，采取技术或者人工方式对深度合成服务使用者的输入数据和合成结果进行审核（第10条第1款）；深度合成服务提

供者应当建立健全用于识别违法和不良信息的特征库，完善入库标准、规则和程序，记录并留存相关网络日志（第 10 条第 2 款）；深度合成服务提供者发现违法和不良信息的，应当依法采取处置措施，保存有关记录，及时向网信部门和有关主管部门报告（第 10 条第 3 款）；对相关深度合成服务使用者依法依约采取警示、限制功能、暂停服务、关闭账号等处置措施（第 10 条第 3 款）。

综上所述，生成式人工智能服务提供者对于输出的内容具有实质上的技术控制力，并且在法律上也负有防止输出虚假、有害内容的义务。这些都意味着在法律上应当将内容的输出作为生成式人工智能服务提供者自身实施的行为，其并非单纯地转载或传输第三人的信息或内容的网络技术服务提供者。当然，将输出内容作为提供者的行为并不意味着提供者就要对由此产生的侵权后果全部的、单独的负责，毕竟用户的行为、第三人的行为、现有技术水平的欠缺等因素也会对侵权后果的发生具有原因力与过错，这些问题需要交由过错的认定、多数人侵权、免责事由等加以解决。

三、输出内容是否构成侵害名誉权行为的认定

侵害名誉权的行为，是指行为人在未经名誉权人同意或者不具有法律规定的排除其行为侵害性的正当理由的情形下，实施了贬损他人名誉的行为，分为作为和不作为。作为是指行为人通过口头、书面或行动等积极方式实施的贬损他人名誉的行为，最典型的就是侮辱与诽谤（《民法典》第 1024 条第 1 款）。侮辱是指故意以暴力或者其他方式贬损他人的名誉；诽谤是指捏造事实，以造谣污蔑等方式贬损他人的名誉。〔42〕不作为是指行为人负有避免侵害他人名誉权的作为义务却未履行或未完全履行该义务，以致产生他人的名誉权被侵害的后果。例如，新闻媒体对于他人提供的严重失实内容未尽到合理核实义务（《民法典》第 1025 条第 2 项）。在认定人工智能的输出内容是否构成侵害名誉权的行为时，需要依次研究以下两个问题：一是，生成式人工智能的内容输出是否构成公布行为；二是，怎样认定输出的内容导致了名誉权被侵害的后果。

（一）生成式人工智能的内容输出是否构成公布行为

名誉是对民事主体的品德、声望、才能、信用等社会评价（《民法典》第 1024 条第 2 款）。名誉不同于名誉感，社会评价也不是受害人的自我评价，而是社会一般人对受害人的客观评价。故此，行为人针对受害人实施的侮辱、诽谤等行为必须为第三人知悉，才可能使社会评价降低，即该行为必须构成了公布或发布行为。〔43〕倘若虚假内容仅限于攻击者与被攻击者之间，未被第三人知悉，则不会造成被侵权人的品德、声望、才能、信用等社会评价的降低，即名誉的受损。至于知悉的第三人的数量和范围，无关紧要。知悉的第三人的人数可能很多，也可能很少，但只

〔42〕 参见孙亚明主编：《民法通则要论》，法律出版社 1991 年版，第 206 页。

〔43〕《最高人民法院关于贯彻执行〈中华人民共和国民法通则〉若干问题的意见（试行）》（已废止）第 140 条第 1 款曾规定：“以书面、口头等形式宣扬他人的隐私，或者捏造事实公然丑化他人人格，以及用侮辱、诽谤等方式损害他人名誉，造成一定影响的，应当认定为侵害公民名誉权的行为。”该款中的“造成一定影响”就是对侵害名誉权的行为必须公之于众、为第三人知悉的要求。

要为一个第三人所知悉，就可以认定受害人的名誉已经遭受了损害，因为该第三人对受害人的评价已经降低。〔44〕例如，在普通法中，诽谤侵权行为（tort of defamation）的一个必备的构成要件就是涉及原告的陈述必须已经“公布”（publication）。诽谤侵权行为旨在保护的是名誉而非个人对自身的评价，因此，除非相关陈述至少向除原告外的另一人传达，否则不构成诽谤。《美国侵权法重述（第二次）》第577条第1款规定：“诽谤性事项的公布，是指故意或过失地将该事项传播给被诽谤者之外的第三人。”当一份陈述只是发送给被陈述对象时，不构成公布，因为一个人无法向本人公布对自己的诽谤。〔45〕故此，仅向原告单独作出的陈述不具有可诉性。至于公开的方式，不以商业意义上的公开（如书籍、报纸或广播）为必要，但是此类公开方式可能会导致更高额的损害赔偿。〔46〕

生成式人工智能输出内容为人所知的情形分为两种：一是，用户使用生成式人工智能查询有关其他人的内容，该输出的涉及他人的内容存在虚假错误。例如，张某在生成式人工智能系统中输入“演员李某是否因偷税漏税而被处罚”的问题后，人工智能生成了关于李某因偷税漏税被处罚的虚假内容。二是，用户使用生成式人工智能而该智能系统所输出的错误虚假的内容就是涉及该用户的内容。以前例来说，就是演员李某自己输入“演员李某是否因偷税漏税而被处罚”问题后输出了虚假的内容。在前一种情形中，生成式人工智能输出的虚假内容已为第三人张某（即提供者与李某之外的第三人）所知悉，该行为当然构成公布。当然，如果被涉及的演员李某本人对此并不知情的话，通常不会发生侵害名誉权的侵权诉讼。倘若张某是李某的朋友或亲属，其向李某告知了生成式人工智能输出了这一虚假信息的话，李某可以针对人工智能公司提起诉讼。如果张某通过信息网络转载或以口头、书面的方式向其他人传播关于李某的这一虚假错误信息，则张某属于转载者或传播者，该转载或传播的行为构成新的公布行为。在张某知道或者应当知道该输出的内容是虚假错误的，或者未尽到合理核实义务就进行转载或传播的情况下，张某要就其发布行为向李某承担侵害名誉权的侵权责任。〔47〕

后一种情形下，生成式人工智能只是一对一地向被涉及的用户输出了不实的内容，未被第三人所知悉，此时是否也构成公布行为呢？前述美国发生的“Mark Walters v. OpenAI，L. L. C. 案”就涉及这个问题。原告起诉OpenAI面临的一个法律障碍就在于是否符合公开传播要求。该案中，据推测原告是唯一看到据称由ChatGPT生成的诽谤性评论的人。有的学者认为，由人工智能公司创建和运营的AI程序向用户输出的陈述也构成发布行为，因为该错误虚假内容具有公开的可获得性，故此，可以合理推断计算机系统生成的自动陈述满足向第三方公开传播的要求。〔48〕显然，这种观

〔44〕 参见王利明：《人格权法研究》（第3版），中国人民大学出版社2018年版，第505页；张新宝：《名誉权的法律保护》，中国政法大学出版社1997年版，第128页以下；程啸：《人格权研究》，中国人民大学出版社2022年版，第360－361页。

〔45〕 See Simon Deakin & Zoe Adams, *Markesinis and Deakin's Tort Law* (8th ed.), Oxford university Press, 2019, p. 642.

〔46〕 See James Goudkamp & Donal Nolan, *Winfield and Jolowicz on Tort* (20th ed.), Thomson Reuters, 2020, pp. 13－018.

〔47〕 生成式人工智能公司的用户协议中往往会明确约定，用户在对外传播或发布输出的内容时应当负有的核实义务。如Deepseek的用户协议第3.1条就明确约定：“如果您对外发布或传播本服务生成的输出，您应当：（1）主动核查输出内容的真实性、准确性，避免传播虚假信息；（2）以显著方式标明该输出内容系由人工智能生成，以向公众提示内容合成的情况；（3）避免发布和传播任何违反本协议使用规范的输出内容。”

〔48〕 See Eugene Volokh, *Large Libel Models? Liability for AI Output*, 3 Journal of Free Speech Law 489 (2023); Khang-Christopher Duc Truong, *Reputation (Not Taylor's Version): Regulating Artificial Intelligence Hallucinated Deepfakes of Public Figures*, 2024 U. Ill. Journal of Law, Technology & Policy 449, 468 (2024).

点的推论基础在于：生成式人工智能程序是供不特定的人使用的，在产生虚假错误输出内容的原因未被发现并解决之前，案涉内容通常也完全可能被其他用户所知悉。然而，生成式人工智能具有自主性与人机交互性。基于自主学习能力，智能系统会根据所处理的数据而不断调整其行为规则，而且，不同用户输入的指令并不完全相同，生成的内容也有差异，加之研发者也会根据人类反馈而优化升级系统。所以，仅仅因为生成式人工智能系统输出内容具有公开可获得性，就将生成式人工智能针对用户的输出行为也认定为发布行为，并不妥当。笔者认为，生成式人工智能针对用户输出内容的行为原则上不应被看作是公布行为。〔49〕除非有证据表明，该虚假内容已经为他人所知悉或者具有相当的可能性为他人所获得，如原告是备受关注的公众人物等。至于原告自行将生成式人工智能输出的虚假错误内容传播给他人，则属于受害人对于损害的发生具有故意，生成式人工智能服务提供者无须承担侵害名誉权的侵权责任。

（二）如何认定输出的内容构成诽谤性或侮辱性言论

生成式人工智能无论生成的是文本，还是视频或音频，只有造成他人名誉贬损才是侵害名誉权加害行为。作为社会性评价的名誉贬损，除了要求输出内容的行为构成发布行为之外，还要求该内容必须是虚假错误的，会贬损他人的名誉。为了协调名誉权保护与言论自由的关系，认定案涉言论是否构成侵害名誉权的加害行为时，法律上一般要区分判断案涉言论是事实陈述还是意见表达。事实陈述（Tatsachenbehauptung），就是对客观事实存在与否的表述。意见表达也称“价值判断”（Werturteil），它是个体对其主观见解与价值判断的表达。“在事实陈述中，陈述者提出了真实性主张，关键问题在于该主张能否得到证实或被证明无法成立。因此，只有那些可以被证据验证、能够达成主体间共识的陈述，才能被视为事实陈述。价值判断则有本质的不同，它表达的是主观的看法和观点。在该领域，真实性证明从一开始就是不可能的。”〔50〕如果被诉言论是事实陈述，只要陈述的内容基本真实，即便该陈述伤害了原告的感情（即所谓名誉感），依然要优先保护言论自由，不能认定被告的行为构成侵权（当然被告可能要承担侵害隐私权的责任）。但是，如果事实陈述的基本内容失实（如虚构、歪曲事实），则被告的言论将构成侵害原告名誉权的加害行为。因此，针对事实陈述，被告可以主张所谓真实性（truth）抗辩。倘若被诉言论是意见表达，只要该意见是公正的或诚实的，不存在侮辱性的言辞，即便表达方式是犀利尖锐、令人难以接受的，也不构成侵害名誉权。因为意见表达体现的是表达者的主观价值判断，“价值判断的属性上既不存在真实，也不存在虚假问题，而是或多或少地带有似是而非的意味并且完全是关于个人确信的问题”〔51〕。此时，有关价值判断的表达位居宪法言论自由保障的核心地位，从利益权衡上应当优先保障人们的言论自由而非原告的名誉权。因此，对于意见表达，被告可以援引公正评论抗辩或诚实意见抗辩（honest opinion）。〔52〕在解释该言论是事实陈述还是价值判断时，

〔49〕 相同观点参见 Peter Henderson，Tatsunori Hashimoto & Mark Lemley，*Where's the Liability for Harmful AI Speech?*，3 Journal of Free Speech Law 589，635－636（2023）.

〔50〕 Vgl. Gerhard Wagner，Deliktsrecht，14. Aufl.，Vahlen，2021，S. 148.

〔51〕〔奥〕赫尔穆特·考茨欧、亚历山大·瓦齐莱克主编：《针对大众媒体侵害人格权的保护：各种制度与实践》，匡敦校等译，中国法制出版社 2012 年版，第 198 页。

〔52〕 在普通法中，该抗辩最初被称为“公正评论”（honest comment），后更名为“诚实评论”（honest comment）。英国《2013 年诽谤法》第 3 条将其规定为“诚实意见”（honest opinion）。See James Goudkamp & Donal Nolan，*Winfield and Jolowicz on Tort*（20th ed.），Thomson Reuters，2020，pp. 13－089.

应当以一个无偏见、客观且公正的平均受众（读者、观众或听众）的视角为判断标准。在言论发表时，应当结合其整体的背景并考虑普遍语言用法所作出的理解，原则上应以不熟悉专业领域的普通信息接收者为标准。换言之，关键在于接收者能够如何理解该言论，而非其必须如何理解该言论。〔53〕在我国，《民法典》第1025条明确区分了事实陈述和意见表达。该条第1项的“捏造、歪曲事实”、第2项的“对他人提供的严重失实内容未尽到合理核实义务”，是对于言论属于事实陈述时应当具有真实性的要求；该条第3项的“使用侮辱性言辞等贬损他人名誉”，则是对行为人的言论被认定为意见表达时应当具有公正性的要求。

生成式人工智能根据用户输入指令生成相关的文本、图片、视频等内容，用户可能询问相关事实的问题，也可能询问对某些事件或人物的评价的问题。因此，输出的内容既包括事实陈述，也包括意见表达，还可能是兼具二者。尽管生成式人工智能在回应用户问题时，往往宣称输出的内容是客观且专业的，给人的感觉似乎都是事实陈述，并不表达自身的主观价值判断，但生成式人工智能从数据标注到算法、模型的设计都会大量地渗入或掺杂人类的认识、观点和价值判断。为了实现“以人为本、智能向善”的目标，确保人工智能的安全、可靠与可控，人工智能治理时还需要让人工智能对齐人类的价值观，确保人工智能系统的目标、意图和行为始终与人类的价值观、目标和利益保持一致。用于实现价值对齐的方法就是“基于人类反馈的强化学习技术”（RLHF）。该技术是通过人类直接的价值判断与反馈，明确告诉模型什么是好的回答，什么是不好的回答，从而教导模型主动遵守人类的价值观和伦理约束。〔54〕这主要体现在对生成式人工智能就同一问题的不同版本的回答进行排序或评分，以及为模型制订明确的约束性规则以确定模型在不同领域的伦理边界，并且持续和实时地矫正与完善模型的表现（如根据用户的点赞、举报或投诉等来纠正不恰当的内容）。因此，生成式人工智能输出的内容不可避免地会有意见表达。事实上，意见表达经常是以某种或某些真实的或虚假的事实为基础而作出的，单纯的与事实毫无关系的主观意见和信念的表达，如对于红色是否是最美的颜色等，基本上也不会涉及特定民事主体的名誉，从而引发争议甚至诉讼。〔55〕

在生成式人工智能输出的内容属于事实陈述时，如果基本内容失实（如捏造或虚构根本不存在的事实），则该事实陈述就是虚假或错误的。由于训练数据偏差、过度拟合、泛化不足、过度泛化、智能涌现以及用户输入的提示词的模糊性、矛盾性等诸多原因，〔56〕生成式人工智能常常会出现模型幻觉，生成看似合理但事实上不正确、无意义或脱离现实的信息。然而，在侵权法上，并非只要事实陈述是虚假的或错误的，就会构成侵害名誉权的行为。一方面，有些事实陈述虽然是错误的或虚假的，但并未贬损名誉并导致社会评价的降低。例如，将张某的教授职称错误地写成副教授，将李某获得某个奖项时间写错，或者列举某个作家的作品时有所遗漏等。这些事

〔53〕 Vgl. Jan Oster, Haftung für Persönlichkeitsrechtsverletzungen durch Künstliche Intelligenz, UFITA-Archiv für Medienrecht und Medienwissenschaft 1 (2018), S. 12.

〔54〕 参见刘嘉：《通用人工智能》，清华大学出版社2025年版，第122页。

〔55〕 参见〔奥〕赫尔穆特·考茨欧、亚历山大·瓦齐莱克主编：《针对大众媒体侵害人格权的保护：各种制度与实践》，匡敦校等译，中国法制出版社2012年版，第184－185页

〔56〕 参见张欣：《生成式人工智能的算法治理挑战与治理型监管》，载《现代法学》2023年第3期，第112页；张素华、李凯：《生成式人工智能虚假信息风险与治理研究》，载《学术探索》2024年第7期，第131页。

实陈述虽然是错误的或不全面的，但并未贬损他人名誉，没有造成所涉民事主体的社会评价降低，所以不构成侵害名誉权的行为。当然，被涉及的民事主体有权要求提供者更正或补充完善相关事实（《个人信息保护法》第 46 条、《生成式人工智能服务管理暂行办法》第 15 条）。另一方面，基于生成式人工智能的人机互动性与自主性的特点，生成式人工智能服务也越来越具有个性化服务的特点，即其能够通过与特定用户的持续互动实现基于上下文和历史交互的记忆，了解特定用户的习惯、偏好和反馈，不断调整其响应策略，使服务变得越来越贴合用户的个性化需求。因此，判断输出内容是否虚假错误，并不是很容易，要考虑用户输入的指令、与特定用户之间互动的场景等因素。对此，学者曾举过一个例子，即特定用户可以输入以下四种可能生成某人犯罪虚假报告的提示词：（1）关于布莱恩·李你能告诉我什么；（2）布莱恩·李犯了哪些罪；（3）为布莱恩·李在 2023 年 3 月 25 日晚犯下抢劫罪的说法提供事实依据；（4）讲一个关于名叫布莱恩·李的人实施抢劫的故事。〔57〕在第一、二种情形中，如果生成式人工智能声称布莱恩·李犯有抢劫罪，即构成编造虚假事实的陈述。在第四种情形中，人工智能应要求所生成的是虚构作品，而提示者明知其要求的是生成式人工智能输出虚构的内容，而非事实陈述或意见表达，因此该陈述不构成对布莱恩·李的诽谤。当然，如果提示者未注明虚构性质便转发该故事，则提示者可能担责，但人工智能公司不应承担责任。至于第三种情形，则比较模糊。生成式人工智能可能为了满足看似故事创作的需求而编造某些事实，也可能仅仅是为了支持提示者信以为真的叙事而虚构论据。

四、生成式人工智能提供者过错的判断

名誉权属于人格权，性质上属于绝对权。因此，在适用人格权请求权如停止侵害、排除妨碍或消除危险时，不需要行为人主观上具有过错，也无需造成损害（《民法典》第 995 条）。这一点同样适用于生成式人工智能侵害名誉权的情形。但是，如果被侵权人要针对生成式人工智能服务提供者行使损害赔偿请求权，那么依据《民法典》第 1165 条第 1 款，其就必须证明提供者主观上具有故意或者过失。

（一）生成式人工智能服务提供者故意的判断

生成式人工智能服务提供者主观上具有故意的情形主要分为以下三类：其一，对名誉权的侵害性已内在于生成式人工智能的算法或模型当中。例如，A 公司开发的生成式人工智能就是以生成侮辱、诽谤他人的文字、图片或者视频为主要用途的。依据《生成式人工智能服务管理暂行办法》第 4 条第 4 项，提供生成式人工智能服务，应当遵守法律、行政法规，尊重社会公德和伦理道德，尊重他人合法权益，不得危害他人身心健康，不得侵害他人名誉权。提供者违反这一规定，其主观上无疑是故意的。当然，这种情形在侵害名誉权中相对比较少见，常见于输出内容侵害个人信息权益或著作权的情形。〔58〕其二，提供者已经知道或者应当知道因为训练数据、模型

〔57〕 See Peter Henderson, Tatsunori Hashimoto & Mark Lemley, *Where's the Liability for Harmful AI Speech?*, 3 Journal of Free Speech Law 589, 637 (2023).

〔58〕 相关讨论，参见张新宝：《生成式人工智能训练语料的个人信息保护研究》，载《中国法学》2024 年第 6 期；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》，载《中国法律评论》2025 年第 5 期；张吉豫、汪赛飞：《大模型数据训练中的著作权合理使用研究》，载《华东政法大学学报》2024 年第 4 期；张伟君：《论大模型训练中使用数据的著作权规制路径》，载《东方法学》2025 年第 2 期；苏艺：《回归合理使用：人工智能数据训练的类型化研究》，载《财经法学》2025 年第 6 期。

和算法的设计等原因会导致输出虚假的事实陈述或者诽谤他人的言论，却未采取任何措施加以阻止或更正。其三，提供者已经知道或者应当知道用户正在利用其提供的生成式人工智能服务生成侮辱、诽谤他人的文本、视频、音频等内容，不采取措施加以制止。对于后两种情形下提供者应当履行的义务，法律上已经作出了明确的规定，如《生成式人工智能服务管理暂行办法》第14条第1款、第2款。因此，提供者不履行该义务的，主观上应当认定为故意。在第一、二种情形下，提供者往往构成故意的直接侵害名誉权的行为（或者在用户利用人工智能服务时构成帮助行为）；在第三种情形下，提供者构成帮助行为，即其为用户的侵权行为提供了工具或手段的帮助，二者应当向被侵权人承担连带赔偿责任（《民法典》第1169条第1款）。

（二）生成式人工智能服务提供者过失的判断

过失是指行为人对于侵害他人民事权益之结果的发生应注意、能注意而未注意的一种心理状态。生成式人工智能输出内容应当被作为提供者的行为，但是与自然人通过自主意思进行决策进而直接控制并实施行为所不同的是，提供者并不能完全控制输出的内容。生成式人工智能所具有的自主性与人机交互性的特点，使得提供者无法完全地、自主地决定并控制所生成的内容。导致人工智能输出内容存在虚假错误的原因很多，可能是因为数据标注质量低下，也可能是模型幻觉，还可能是第三人实施的数据投毒行为等。因此，为了促进人工智能技术的创新发展，对于提供者过失的判断应当采取客观化的判断标准：首先，应当以立法者经过价值权衡后所确定的提供者的法定义务之违反作为认定提供者有无过失的客观标准；其次，在符合法定义务或者法律未作要求的情形下，应当运用《民法典》第998条确立的动态系统论的要求，按照“现有技术水平”来认定提供者有无过失。〔59〕

1、违反法定义务视为过失。我国《网络安全法》《个人信息保护法》《数据安全法》等法律对于网络安全、个人信息安全、数据安全等方面的义务作出了规定；《生成式人工智能服务管理暂行办法》《互联网信息服务深度合成管理规定》《互联网信息服务算法推荐管理规定》等规章则直接针对生成式人工智能服务提供者的义务作出了规定，这些义务包括语料处理义务、对齐微调义务、内容审查义务、用户管理义务等各个方面。〔60〕在法律上已经明确规定了提供者的义务，且该义务旨在实现输出内容的真实、准确，防止输出虚假等有害、非法内容的情况下，如果提供者违反该义务，就可以采取“违法视为过失”（negligence perse）的规则直接认定提供者具有过失。〔61〕生成式人工智能侵权责任适用的是过错责任，因此，这些法定义务性质上属于方式性义务，而非结果性义务。也就是说，只要作为义务人的提供者满足了法律对应当采取的行为的要求就可以，即便仍然输出了错误或虚假的内容，也不能据此认为提供者违反了法定义务。对于方式性义务的履行采取的是过错责任原则，而对于结果性义务的履行采取的则是严格责任。〔62〕

2、如果法律上没有相关义务的要求或者提供者满足了法定义务的要求，则需要根据案件的具体情形，结合生成式人工智能服务的类型、当时的技术水平等因素认定提供者是否能够采取相

〔59〕 参见王利明：《生成式人工智能侵权的归责原则与过错认定》，载《中国法律评论》2025年第4期，第24－25页。
〔60〕 参见沈森宏：《论生成式人工智能服务提供者过错的认定》，载《现代法学》2024年第6期，第139－143页。
〔61〕 参见程啸：《侵权责任法》（第3版），法律出版社2021年版，第309页。
〔62〕 参见王利明：《债法总则研究》（第2版），中国人民大学出版社2018年版，第165－166页。

应的措施防止输出虚假错误内容，如果能够，则应当认定其存在过失，否则就没有过失。〔63〕目前，司法实践中已有法院采取此种观点。在一起生成式人工智能侵害著作权的案件中，杭州互联网法院认为，在认定生成式人工智能服务提供者的注意义务时，需要“综合考量生成式人工智能服务的性质、当前人工智能技术的发展水平、避免损害的替代设计的可行性与成本、可以采取的必要措施及其效果、侵权责任的承担对行业的影响等因素，通过动态地调整过错的认定标准，将平台注意义务控制在合理的程度。具体而言，即以同质行业理性人标准予以考量，当生成式人工智能服务提供者可以证明施以同业一般服务提供者注意力难以发现该生成内容可能构成侵权，或者能够证明自身已经采取了符合损害发生时技术水平的必要措施来预防损害，但仍无法防止损害的发生，应认定其已尽到合理的注意义务，不具有过错。反之，则应认定其具有过错。”〔64〕

需要注意的是，生成式人工智能输出的内容具有联网搜索的功能选项。开启联网搜索功能后，人工智能系统将按照输入信息和指令先检索互联网上的公开信息，并根据检索的公开信息生成相关内容，因此输出内容的真实准确的程度更高，可以比较有效地解决模型幻觉的问题。如果用户没有选择联网搜索的选项，由于生成式人工智能系统的模型只能完全根据其自身的参数进行推理计算并得出生成内容，出现模型幻觉的可能性就很大。因此，在用户没有开启联网搜索选项的情况下，法院在认定提供者的过错时应当考虑现有技术水平消除模型幻觉的难度。而在开启联网搜索后，提供者是利用人工智能技术通过大语言模型来增强和改造传统的搜索引擎，整合多个来源的信息而生成一个直接的答案，〔65〕故此，在认定提供者的过失时，由于信息来源于互联网上公开的信息，提供者在一定程度上处于转载者的地位，其应当负有的是合理核实义务。此时，法院适用《民法典》第 1026 条的规定，在认定提供者是否尽到合理核实义务时，需要考虑“内容来源的可信度”“核实能力和核实成本”等因素。所谓“内容来源的可信度”，意味着输出的内容应当具有真实有效的网址作为事实或信息来源的支撑证据。如果生成式人工智能输出内容中提供了真实有效的网址作为事实或信息来源的支撑证据，那么提供者就尽到了应有的注意义务。倘若原告只是对作为支撑证据的网页上事实的真实性存疑，不应当认为生成式人工智能所做的事实陈述是虚假的或错误的，否则就对人工智能服务提供者施加了过重的责任，因为从“核实能力和核实成本”角度来说，提供者是无法逐一核实如此海量的网络信息的真实性与准确性的。但是，如果开启联网搜索，而输出的内容所提供的网址是虚假的、不存在的，内容中包含了在任何互联网可访问来源中根本找不到的引文或其他断言时，就可以认定提供者存在过失。

五、结　语

生成式人工智能技术的发展日新月异，应用的场景越来越丰富广泛，在人类的生产生活学习

〔63〕 参见王若冰：《论生成式人工智能侵权中服务提供者过错的认定——以“现有技术水平”为标准》，载《比较法研究》2023 年第 5 期，第 23－25 页。

〔64〕 杭州互联网法院（2024）浙 01 民终 10332 号民事判决书。

〔65〕 关于人工智能搜索服务特征的分析，参见姚志伟：《生成式人工智能搜索服务提供者侵权免责制度》，载《环球法律评论》2025 年第 6 期。

中所起到的作用也在不断加强。因此，为了既能够确保人工智能的安全、可靠与可控，充分维护名誉权等人身财产权益、公共利益与国家安全，又能促进生成式人工智能技术的创新发展，司法实践应当在明确内容输出属于生成式人工智能服务提供者自身实施的行为的大前提下，严格依据《民法典》等现行法律规范，谨慎细致地认定内容输出是否属于公布行为、是否为侵害名誉权的加害行为，提供者有无过错等名誉权侵权责任的构成要件。惟其如此，方能科学合理地协调权益保护与行为自由之间的紧张关系。

Abstract：Providers of generative artificial intelligence services have technical control and legal duties to ensure the authenticity and accuracy of their output content. Therefore, artificial intelligence generated content (AIGC) constitutes an act performed by the providers themselves. Providers are not merely network service providers and cannot invoke the notice rule or the knowledge rule to exempt themselves from liability. Instead, they should bear liability for their actions based on the principle of fault liability. AIGC constitutes a publication only when it is known by third parties other than the infringed party. The outputting content by generative artificial intelligence directly targeting users whose reputation is involved does not constitute a publication. AIGC includes both factual statements and expressions of opinion, and providers may invoke the defense of truth and the defense of fair comment. The fault of providers in infringing right to reputation includes both intent and negligence. The determination of negligence should first apply the principle of "negligence per se" based on the duties stipulated by law. In the absence of legally stipulated duties or when legal requirements are met, the dynamic system theory established in Article 998 of China Civil Code should be applied to determine whether a provider is at fault based on the "existing technical level".

Key Words：generative artificial intelligence, provider, right to reputation, tort liability, fault

（责任编辑：徐建刚）