Dispute Settlement 争议解决, 2026, 12(4), 100-111
Published Online April 2026 in Hans. https://www.hanspub.org/journal/ds
https://doi.org/10.12677/ds.2026.124106

Hans汉斯

论网络暴力治理中平台责任与刑事归责的联动机制

朱晓双

西南民族大学法学院，四川 成都

收稿日期：2026年3月2日；录用日期：2026年3月26日；发布日期：2026年4月7日

摘　要

数字时代，网络暴力不仅侵犯个人权益，而且破坏健康的生态网络，已成为严峻的社会治理难题。“杭州取快递被造谣案”生动揭示了传统以个人为中心的刑法规制模式，在面对涉众型网络暴力时存在“法不责众”、主观故意难认定、因果关系模糊等现实困境。与此同时，数字平台所具有的网络空间枢纽地位、技术优势与“私权力”属性，在网络暴力的生成、发酵与扩散过程中扮演着关键角色，却常因“避风港”原则而在治理中趋于消极。本文认为，破解网络暴力治理困局的关键在于构建平台责任与刑事归责的联动机制。首先，应在理论上重塑平台角色，将其定位为积极的“网络看门人”，课以体系化的信息网络安全管理义务。在此基础上，构建双向联动路径：一方面，平台积极履行事前预防、事中干预、事后救济等义务的行为，可以作为减免其刑事责任的依据，实现正向激励；另一方面，当平台“明知”或“应知”网暴信息存在而消极不作为，造成严重后果时，应通过激活拒不履行信息网络安全管理义务罪、追究其片面帮助犯的刑事责任等方式实现反向规制。通过这种双向互动，最终形成“国家管平台、平台管用户”的治理新范式，实现网络暴力治理中效率与公正的统一。

关键词

数字平台，网络暴力，平台责任，刑事归责，联动机制

On the Linkage Mechanism between Platform Responsibility and Criminal Imputation in the Governance of Cyber Violence

Xiaoshuang Zhu

Law School, Southwest Minzu University, Chengdu Sichuan

Received: March 2, 2026; accepted: March 26, 2026; published: April 7, 2026

文章引用: 朱晓双. 论网络暴力治理中平台责任与刑事归责的联动机制[J]. 争议解决, 2026, 12(4): 100-111.
DOI: 10.12677/ds.2026.124106

Abstract

In the digital era, cyber violence has emerged as a formidable governance challenge, infringing upon individual rights and undermining the integrity of the online ecosystem. The case of "Defamation Against a Woman Picking Up a Package in Hangzhou" exposes the limitations of traditional criminal law, which centers on individual liability, when addressing collective cyber violence. These limitations include difficulties in establishing subjective intent, proving causation, and the dilemma of "impunity for the masses". Digital platforms, by virtue of their pivotal position, technical capabilities, and exercise of "private power", profoundly influence the evolution of cyber violence; yet, under the "Safe Harbor Principle", they often remain passive. This paper argues that an effective governance mechanism necessitates the integration of platform liability with criminal imputation. Platforms should be repositioned as proactive "gatekeepers" with clearly defined obligations to ensure information security. Furthermore, a bidirectional linkage mechanism should be constructed: on one hand, platforms that actively fulfill their duties of prevention, intervention, and remediation should receive positive incentives; on the other hand, platforms that knowingly or negligently fail to fulfill their cybersecurity management obligations despite being aware of ongoing cyber violence, thereby causing serious consequences, should face criminal liability either under the crime of refusing to fulfill information network security management obligations or as accomplices via unilateral complicity. This bidirectional approach aims to establish a governance model of "the state regulates platforms, and platforms regulate users", thereby achieving both efficiency and justice in the governance of cyber violence.

Keywords

Digital Platforms, Cyber Violence, Platform Liability, Criminal Imputation, Linkage Mechanism

1. 引言

随着互联网技术对社会结构的深度嵌入，网络空间已成为与现实社会交互并行的“第二生活空间”。然而，技术赋能的同时也带来了治理挑战，网络暴力如“按键伤人”甚至“按键杀人”等现象频发，严重侵害公民人格权益，破坏网络生态。从德阳女医生案[1]、武汉小学生母亲坠楼案[2]、粉色头发女孩郑某华自杀案[3]、寻亲男孩刘学州自杀案[4]等系列事件中看出，网络暴力导致悲剧不断上演，引发了全社会对网络暴力治理的深刻反思。

在此背景下，国家层面密集出台了一系列规范文件，如《关于切实加强网络暴力治理的通知》《网络暴力信息治理规定》，以及《关于依法惩治网络暴力违法犯罪的指导意见》(以下简称《指导意见》)等，力图构建全方位的治理体系。其中，一个转变日益清晰：治理的重心正从单纯惩罚具体的“个人施暴者”，转向强化作为关键节点的“数字平台”的责任[1]。

[1]参见四川省德阳市中级人民法院刑事判决书，(2021)川 06 刑终 125 号。
[2]参见键盘上的风暴：《母痛未息，网暴再伤》，微信公众号，2025 年 4 月 29 日 https://mp.weixin.qq.com/s/7M1lFIo0v3hg5yf1Qnc16w。
[3]参见胡欣红：《“粉发女孩”离世网暴悲剧何时终结？》，载《工人日报》2023 年 2 月 22 日，第 5 版。
[4]参见赵丽、杨轶男：《河北寻亲男孩轻生网暴被指是重要诱因“键盘侠们的狂欢”该结束了》，载《法治日报》2022 年 1 月 26 日，第 4 版。

然而，当前的理论研究与司法实践，对于如何精准界定平台责任，尤其是如何将平台责任与传统的刑事归责体系有效衔接，仍存在诸多模糊地带。一方面，刑法对网络暴力行为的规制，主要聚焦于侮辱、诽谤、侵犯公民个人信息等具体罪名，面对参与者聚合、主观状态复杂、因果关系模糊的涉众型网络暴力，常陷入“法不责众”的困境[2]。另一方面，平台的责任边界也常在“避风港”原则[5]下的被动角色与作为“守门人”的积极义务之间摇摆，导致其在网暴治理中或动力不足，或手段单一，难以发挥应有作用[3]。

“杭州取快递被造谣案”即郎某、何某诽谤案[6]正是这一系列矛盾的集中体现。该案从自诉到公诉的程序转换，展现了司法机关能动回应网络时代挑战的努力，但也深刻暴露了被害人维权的艰辛与传统追诉模式的局限。本文将以该案为分析起点，聚焦于一个核心问题：在数字时代，应如何构建平台责任与刑事归责的有效联动机制，使平台从被动的信息“容器”转变为主动的风险“守门人”，从而实现网络暴力治理的根本性突破。本文首先剖析传统规制路径的失灵与平台治理的困境，继而论证平台角色重塑的理论基础与义务体系，最后提出构建平台责任与刑事归责双向联动的具体路径，以此寻求网络暴力治理的可行性方案。

2. 二元困境：传统规制路径的失灵与平台治理角色的缺位

网络暴力治理的困境并非单一维度的法律适用难题，而是植根于数字时代社会治理结构的深层巨变。传统刑法以“行为人”为中心的归责逻辑，与网络暴力“涉众型”“累积性”的犯罪形态之间形成结构性错位[4]；而作为网络空间实际掌控者的数字平台，却在“避风港”原则的庇护下长期扮演消极角色。2020 年发生的“杭州取快递被造谣案”，是网络暴力治理史上的一个里程碑式案件。它不仅因情节恶劣、后果严重而备受关注，更因其诉讼程序的戏剧性转折，从行政处罚到刑事自诉，再到检察机关主动介入建议转为公诉，深刻揭示了现有法律框架在面对新型网络暴力时的结构性困境。

(一) 案件聚焦：杭州取快递被造谣案

2020 年 7 月，一起看似普通的诽谤事件，最终演变为推动网络暴力治理法治进程的标志性案件。

1) 基本案情

2020 年 7 月 7 日 18 时许，被告人郎某在浙江省杭州市余杭区某小区快递驿站内，使用手机偷拍正在等待取快递的被害人谷某，并将视频发布在某微信群。随后，郎某与何某合谋，分别假扮快递员和谷某，捏造了谷某结识快递员并多次发生不正当性关系的微信聊天记录。为增强可信度，二人还捏造了“赴约途中”“约会现场”等视频、图片。7 月 7 日至 16 日，郎某将上述捏造的聊天记录截图 39 张及视频、图片陆续发布在该微信群，引发群内大量低俗、侮辱性评论。

同年 8 月 5 日，上述偷拍的视频以及捏造的微信聊天记录截图 27 张被他人合并转发，相继扩散到 110 余个微信群，群成员约 2.6 万人；7 个微信公众号，阅读数达 2 万余次；及 1 个网站，浏览量超 1000 次，其他网络平台也陆续转发。事态持续发酵，至 8 月至 12 月，此事经多家媒体报道引发网络热议，仅微博话题“被造谣出轨女子至今找不到工作”阅读量就达 4.7 亿次、话题讨论 5.8 万人次。被害人谷某因此被公司劝退，多次求职被拒，遭受严重经济损失和精神创伤。

本案经过行政处罚到被害人自诉再到检察机关提起公诉，从一审到二审，多级程序流转。最终法院

[5]“避风港”原则起源于美国 1998 年制定的《数字千年版权法案》(DMCA, Digital Millennium Copyright Act)是指法律为网络服务提供者如视频网站、电商平台、网盘等设立的一个“免责港湾”。核心内容是：当平台上的用户未经许可上传了侵权内容(如盗版视频、假冒商品图片)时，只要平台本身不知道侵权内容的存在，并且在接到权利人的合格通知后，迅速删除了侵权内容或断开了链接，那么平台就可以不承担侵权赔偿责任。我国在立法时借鉴了这一制度，将其本土化并融入到了多部法律法规中，形成了一个完整的体系。

[6]参见浙江省杭州市余杭区(市)人民法院：(2021)浙 0110 刑初 180 号。

以行为人通过网络对不特定对象实施诽谤，且诽谤信息在网络上大范围流传，引发大量淫秽、低俗评论，造成不特定公众恐慌和社会安全感、秩序感下降，引发网络秩序混乱，认定其诽谤行为已严重危害社会秩序，按照刑法第二百四十六条第二款的规定处理。

2) 案件进程引发法理争议与治理反思

本案的程序流转极具典型意义，折射出网络暴力治理中的深层法律困境。从被害人谷某 8 月 7 日报案后公安机关仅作行政拘留处罚，到 10 月 26 日法院受理刑事自诉，再到 12 月 22 日检察机关以“严重危害社会秩序”为由建议转为公诉，这一由“行政–自诉–公诉”构成的程序跃迁，引发了学界的广泛热议。

争议的焦点在于：针对特定个人的诽谤行为，如何突破“告诉才处理”的原则，启动国家公权力追诉。检察机关给出的理由是“被侵害对象系随意选取，具有不特定性，任何人都可能成为被侵害对象，严重破坏了广大公众安全感”。由此，网络暴力的危害已从“个体法益”溢出至“社会法益”，传统依赖被害人意志的追诉模式在网络空间的适用性提出了严峻挑战[4] (pp. 51-54)。

然而，比追诉模式转型更值得深思的是平台责任的缺位[5]。从信息首次发布到最终引爆全网的关键窗口期内，平台几乎处于“隐身”状态，既未主动识别、拦截虚假信息，也未及时采取限流、屏蔽等措施。这种治理真空不仅为诽谤信息的指数级扩散提供了技术温床，也直接导致后续刑事追诉程序的复杂化与被动化。

(二) 传统个体刑事归责局限

在传统刑法理论中，侮辱罪、诽谤罪等被视为“点对点”的犯罪，其归责逻辑建立在明确的“行为人–被害人”二元关系之上[6] (p. 135)。然而，当这一逻辑被移植到网络空间时，却难以适用。

1) “法不责众”的归责难题

首先，是“法不责众”的归责难题。网络暴力的典型特征是“涉众性”。在“杭州案”中，郎某、何某是诽谤信息的“生产者”，但造成被害人谷某“社会性死亡”的真正力量，是后续成千上万的转发者、评论者和围观者，但个体责任的认定存在一定障碍。一方面，网络暴力危害后果是由群体行为以及各种因素共同作用的结果。另一方面，我国刑法存在罪量要素，单个个体行为无法到达此要求[4] (pp. 51-54)。网络空间的切割力使得完整的犯罪行为被分割为海量的、看似危害轻微的“微行为”[7]。刑法若要追究每一个转发者的责任，既无现实可能性，也违背了刑法的谦抑性原则。但若仅处罚发起者，又无法涵盖网络暴力真正的危害来源，导致“雪崩时没有一片雪花认为自己有责”的困局。

2) 主观要件的认定难题

其次，是主观要件的认定难题。网络暴力参与者心态复杂，既有“正义感”驱使下的道德审判，也有盲目跟风的情绪宣泄，更有少数人的恶意中伤。网络媒体的匿名特征让某些网民解除道德自律，有些网民辨别能力不足，站在偏激立场不断输出带有道德偏见的谴责，最终沦为网络暴力参与者。但道德失序无法界定为法律上犯罪故意[8] (p. 86)。要准确界定行为人的主观“明知”或“故意”，在实践中几乎不可能。正如“杭州案”中，大量网民是在信息不完整、甚至被误导的情况下发表评论，其言论虽具攻击性，却难以被证明具有刑法意义上的诽谤故意。网络加害者往往依赖于数据控制者所塑造的场域，其认知容易受到算法影响而产生偏差，刑法若仍按传统路径评价，必然产生谬误[9] (p. 87)。

3) 因果关系与结果归属的难题

最后，是因果关系与结果归属的难题。网络暴力的危害结果是“累积”性的[10]。个体的转发、评论行为与被害人最终的精神崩溃乃至自杀之间，并不存在直接的、物理性的因果链条。这种“多因一果”的复杂局面，使得传统相当因果关系理论难以适用。虽然在司法解释中，将“造成被害人自杀”等严重后果视为“情节严重”的一种表现，但这更多是一种政策性的“缓和的结果归属”[11] [12]，并未真正解

决将整体危害归责于个体行为人的理论障碍[13]。例如，在“杭州女子取快递被造谣案”中，诽谤者、传播者，平台等多方主体共同导致被害人的社会评价严重贬损，属于“多因一果”，但限于现行共犯理论与因果关系规则，只能追究初始诽谤者的刑事责任[6] (p. 140)。这种“个体归责”处理方式显然无法全面评价网络暴力行为的社会危害性，亦不能满足风险社会背景下刑法预防功能的实现需求。

(三) 平台治理角色缺位

当传统刑法对个体施暴者束手无策时，人们自然将目光投向承载这一切的平台。然而，既有法律框架下的平台责任设计，同样存在根本性缺陷。

1) “通知－删除”规则的时效困境

以“避风港”原则为核心的“通知-删除”规则，预设了一个“技术中立”的平台角色即平台是被动的信息存储通道，只有在接到权利人通知后，才负有删除义务。这种事后、被动的责任模式，在面对高速裂变传播的网络暴力时，显得尤为迟钝。在“杭州案”中，从郎某于 7 月 7 日首次发布信息到 8 月 5 日信息被大规模转发引爆舆论，整整经历了近一个月。在此期间，平台既未主动识别异常传播的虚假信息，也未及时对微信群内大量低俗评论进行干预。当被害人最终提起维权时，损害后果已经形成，删除已无法“恢复原状”。平台治理的滞后性，使其在法律程序中处于“事后补救者”的尴尬位置。当然，我国《刑法》第 286 条规定了拒不履行信息网络安全管理义务罪，若网络平台明知存在网络暴力行为时应履行“通知－删除”的义务，但由于“避风港”导致条款沦为僵尸条款，适用率极低[14]。24 年的《网络暴力信息治理规定》虽然强调了网络平台“守门人”的角色，实施从政府到平台再到个人的多元协同治理模式，但如何让网络平台从消极中立转向积极履责，还有需要具体法规加强对网络平台的监管。

2) 流量逻辑与公共责任的冲突

此外，更深层的问题在于，数字平台早已不是“中立”的技术中介。它们凭借算法推荐、流量分配、社群规则制定等权力，深度参与了网络内容的“生产”和“传播”。数字平台通过助推特定议题、加剧失真信息扩散、强化单向度道德审判，实际上已成为网络暴力这一“平台现象”的“隐形推手”[15] (pp. 129-130)。为了追求商业利益和用户粘性，算法逻辑往往优先推荐具有争议性、情绪性的内容，客观上为网络暴力的滋生提供了“温床”[16]。“杭州案”中，诽谤信息之所以能从一个小微信群扩散到全网，最终引爆 4.7 亿阅读量的舆论风暴，离不开平台推荐机制对敏感话题的放大效应。平台在追求“流量红利”的同时，却拒绝为流量带来的负面后果承担责任，这种“收益内部化、成本外部化”的商业模式，构成了治理困境的深层根源。

3) “私权力”与“公责任”的结构性错位

除了前述问题，网络平台治理还存在“私权力”与“公共责任”的结构性错位，使平台陷入了深刻的角色冲突之中。作为拥有巨大技术能力和数据权力的“私权力”主体，它具备成为高效“守门人”的潜力；数字平台凭借其对数据、算法和核心用户流量的垄断性控制，对平台内的亿万用户行使着事实上的管理权、奖惩权和纠纷裁决权。这种“私权力”虽源自私人契约，但其效果却具有公共性[9] (pp. 90-96)。平台基于数据垄断地位获得了维护网络秩序的“准公权力”，但作为追求利润的商业主体，它又缺乏主动筛查、干预内容的动力，甚至存在“用网暴换流量”的逆向激励。且现有法律框架并未将前述权力与相应的公共责任有效挂钩。当平台既不积极履行“看门”职责，而传统的个人追责路径又已失灵时，网络暴力的治理便出现了一个巨大的“责任真空”。此时，若不能准确定位平台责任义务，既会削弱网络治理的效果，也会产生监管漏洞，使平台失范行为难以追责。

综上，前述几种困境，核心问题在于现有法律法规多仅作原则性、框架性的规定，既未对平台设定具体可行的内容管理义务，亦未明确相应的法律责任，导致各类平台对于自己领域中的违法信息，普遍存在“不便管、不能管、不愿管”的客观痛点与主观惰性[17]。若仅进行传统个人刑事归责或仅单方面强

化平台责任，忽视其实际管控能力与合理利益诉求，不利于平台网络治理角色的积极转变，难以遏制网络暴力。

3. 嬗变与证成：平台角色重塑的理论基础与义务体系

破解上述困境，必须超越“个人中心主义”的传统刑法观，正视平台在社会治理中的结构性地位，从法理上确立其承担更积极、更前置法律责任的正当性。

(一) 平台角色的嬗变

传统的“避风港”原则与“技术中立”原则，其逻辑前提是将平台视为一个与用户无异的、平等的信息传播参与者，然而数字平台经济的蓬勃发展，已经彻底改变了这一现状。平台凭借其对数据资源、算法技术和核心用户流量的垄断性控制，对平台内的亿万用户行使着事实上的掌控权，获得“拟政府化”的控制地位[18]。这种权力虽源自私人契约，但其效果却具有公共性，因此被学界形象地称为“私权力”[9] (pp. 90-96)。平台可以决定谁在平台上发言、什么样的内容可以被看见、谁可以被“禁言”甚至“驱逐”。这种强大的社会控制力，使得平台具备了“准政府”的特征[19]。相应地，它也必须承担与这种权力相匹配的公共责任。

(二) 平台义务的归责基础

网络暴力的生成与传播机制，进一步强化了平台的归责基础。如前所述，网络暴力的大规模、高强度伤害，高度依赖平台的聚合与放大效应。如果没有平台的算法推荐、热搜榜单、社群串联等功能，个体的恶意言论将永远处于孤立的、低杀伤力的状态。

但平台的技术架构和规则设计，实际上塑造了网络暴力的风险环境。根据危险源控制理论，谁创造了风险，或者谁更有能力以更低成本控制风险，谁就应当承担安全保障义务。平台作为网络空间这一危险源的“开启者”和“支配者”，具有监管危险、防止损害发生的“监管保证人”地位[20]。这种基于排他性技术控制的保证人地位，是其承担不作为刑事责任的法理根基。所谓保证人地位，是指行为人基于其对结果发生原因的控制支配，而在法律上负有防止结果发生的义务[21]。尽管“技术中立”原则常被援引为抗辩理由，但这并不能成为其豁免责任的绝对依据。平台在享受流量红利与用户活跃度收益的同时，也应承担伴随的风险治理成本。其不作为即不删除、不屏蔽与网暴损害结果的扩大之间存在间接的因果联系[19]。因此，正是基于其对信息传播技术与规则的实质性掌控，网络平台应当承担起与其技术能力相匹配的监管保证人责任，成为平台治理的归责基础。

(三) 平台义务体系的来源

平台在网络暴力治理中的义务并非凭空创设，而是有着坚实的法律渊源。

在宪法层面，网络平台具有基本权利保护义务。我国《宪法》虽未直接规定平台义务，但国家保护公民人格尊严、通信秘密等基本权利的义务，需要通过立法传递给包括平台在内的社会权力主体[8] (pp. 152-153)。平台在行使“私权力”时，必须尊重和保障用户的基本权利，不得滥用其支配地位。这构成了平台承担治理责任的宪法基础。

在民法层面，平台还需要履行安全保障与信义义务。《民法典》规定了网络服务提供者的侵权责任，特别是第 1198 条的安全保障义务，为平台保护用户免受第三人侵害提供了民法依据。网络平台为公众提供了一个具有社交属性的公共场所，理应确保该场所的交往安全，防范来自第三人的不法侵害。这意味着平台需要采取合理措施，保护其用户免受网络暴力的侵扰。此外，用户与平台之间存在一种“委托-代理”关系。用户的数字生活、数据安全乃至精神安宁，都高度依赖于平台的行为。作为受托人，平台负有忠实于用户利益、勤勉尽责的信义义务[22]。当平台为追求流量而放任网络暴力损害用户权益时，便构成了对这一根本义务的违反。

在公法层面，平台拥有强制性的信息网络安全管理义务。这是由《网络安全法》《个人信息保护法》以及《网络暴力信息治理规定》等公法规范明确规定平台的强制性义务。《网络暴力信息治理规定》第七条明确要求网络信息服务提供者“履行网络信息内容管理主体责任，建立完善网络暴力信息治理机制，健全用户注册、账号管理、个人信息保护、信息发布审核、监测预警、识别处置等制度”。这一规定将平台义务从宏观的责任细化为覆盖信息传播生命周期的具体合规要求[23]，使平台义务从“软法”变成了具有强制力的“硬约束”[24]。《指导意见》第六条将违反此类义务，经监管部门责令改正而拒不改正，造成严重后果的行为，纳入拒不履行信息网络安全管理义务罪的规制范围，更是将行政法义务与刑事责任直接挂钩，形成了从行政规制到刑事制裁的完整责任链条[4] (pp. 60-64)。

4. 正向激励：平台内部治理义务的体系化构建

网络平台在当下治理角色的转变以及其拥有的信息的传播和技术的控制性地位，决定其须承担更多的公共责任。要平台从消极被动的“技术中立”转向积极的“守门人”角色，则更需要强调平台的刑事责任，尤其要构建好平台治理与刑事归责的联动机制。而构建联动机制的第一步，是明确并强化平台自身的治理义务，将“软法”变成“硬约束”，用积极的义务履行换取刑事责任的豁免，实现正向激励。这种正向激励的核心在于，将平台内部治理的有效性，作为其减免刑事责任的法定事由，从而将国家的治理目标内化为平台的商业自觉。本文认为，要构建此种正向激励机制可分别从事前、事中进行技术治理，在辅以事后法律救济三个维度，细化和体系化平台治理，有效遏制网络暴力[25]。

(一) 事前预警义务

为落实“抓前端、治未病”理念，网络暴力治理的刑事归责应当强调犯罪的前端治理和源头预防。相较于法律治理手段的滞后性，平台的技术治理在网络暴力事件的前期发酵阶段和规模扩大阶段发挥着重要作用，能够有效减少或避免网络暴力危害后果的产生[4] (pp. 60-64)。因此事前和事中的技术治理优于事后的法律救济[26]。网络暴力往往呈现裂变式扩散特征，有效治理网络暴力要求平台以一个更加积极主动、未雨绸缪的姿态，优化网络争议事件的讨论框架，将网络舆情极化消弭在事前，将网络暴力风险扼杀于萌芽。具体而言，平台不能仅满足于“接到通知后删除”。根据《网络暴力信息治理规定》第十二条和第十三条的要求，平台必须建立网络暴力信息识别模型和预警机制[7]。这意味着平台需要综合事件类别、针对主体、参与人数、信息内容、发布频次等维度，利用大数据和人工智能技术，对潜在的网暴风险进行动态监测和早期预警。

在用户注册阶段，平台应当通过社区公约和用户协议明确告知禁止的网络行为。同时，平台还可以通过信息内容、讨论板块、参与人数等维度，对热点事件提前研判，防止舆论激化和情绪极端化的情况出现。例如，新浪微博设立“反网络暴力”超话专区，配置专职巡查员，支持公众人物管理粉丝社群；并通过“评论罗伯特”社交机器人引导评论氛围，以正面信息对冲恶意评论，实现风险预防[14]。此外，当一个普通用户的账号在短时间内涌入大量带有侮辱性关键词的评论或私信时，平台系统应能自动识别并将其判定为高风险事件，立即启动应急响应。这种“算法对抗算法”的技术治理路径，是实现源头防范的关键[15]。抖音等平台已上线“发文警示”功能，对检测到的可能发布不当言论的用户进行提示，便是主动感知风险的有益尝试[27]。平台若已建立并有效运行了符合行业标准的预警模型，即便偶有网暴信息未被即时拦截，也应成为判断其已尽到合理注意义务、减轻其行政或刑事责任的考量因素[15]。

[7]《网络暴力信息治理规定》第十二条规定：网络信息服务提供者应当在国家网信部门和国务院有关部门指导下细化网络暴力信息分类标准规则，建立健全网络暴力信息特征库和典型案例样本库，采用人工智能、大数据等技术手段和人工审核相结合的方式加强对网络暴力信息的识别监测。第十三条规定：网络信息服务提供者应当建立健全网络暴力信息预警模型，综合事件类别、针对主体、参与人数、信息内容、发布频次、环节场景、举报投诉等因素，及时发现预警网络暴力信息风险。……

(二) 事中干预义务

在网络暴力的规模扩大期，需要平台精确定位并实时处置信息传播的关键舆情节点，有效实现事中阻断，及时阻止事态的扩大。在这一阶段，“避风港”原则不应成为平台不作为的借口。在网暴治理领域，应引入并扩大解释“红旗原则”[8]。当网络暴力信息已“像红旗一样飘扬”，达到任何一个理性人都能认识到其违法性的程度时，平台便不能再主张“不知道”，而应当主动采取干预措施。这种“明知–屏蔽”的动态义务包括但不限于：

1) 内容限流与折叠：对已被识别为争议性、煽动性的内容，降低其在信息流、热搜榜的推荐优先级，避免其持续获得流量加持而极端化。

2) 评论管控与风险提示：对明显含有侮辱、诽谤、人身攻击的评论进行自动屏蔽或折叠处理，防止其出现在公共可见区域[23]。此外，平台还可以建立舆情关联话题管理机制，及时发现并切断网暴舆情与官民对立、性别歧视等深层社会矛盾议题的二次关联，防止网络暴力信息的“搭车”传播导致危害升级[28]。对已被证实为谣言或具有高度误导性的信息，主动添加“争议信息”“已被辟谣”等风险提示标签，引导用户理性判断。

3) 紧急防护功能：优化“一键防护”模式，允许用户在遭受攻击时，一键关闭所有陌生人评论、私信和转发功能。这种功能应平等地面向所有用户，而非仅限于付费会员[4] [26] (pp. 60-64)。《网络暴力信息治理规定》第二十三条明确规定，网络信息服务提供者应当提供便利用户设置屏蔽陌生用户、禁止转载或评论本人发布信息等防护选项[25]。若在“粉发女孩郑某华案”中，平台能在海量负面评论涌现之初，自动或应其请求及时启动高级别防护，或许能为她提供一道至关重要的心理防线。

(三) 事后救济义务

在事后救济阶段，损害发生后，平台的义务不应止于删除信息，更重要的，是协助受害者进行有效的法律维权。具体表现为信息备份义务和协助取证义务。

首先，平台应履行好信息备份义务。《网络安全法》第二十一条规定了平台采取数据分类、重要数据备份和加密等措施，并按照规定留存相关网络日志的义务[29]。在司法实践中，电子数据相较于传统证据具有更强的证明力，但在现实层面，网络暴力的被害人时常面临证据难收集、真实身份难核实、个人权利难维护难等诸多难题。原因在于网络暴力的证据材料通常需要执法部门信息溯源和程序流转，取证技术门槛高、难度大，且易灭失或被篡改。此时，平台应当利用自身的数据资源和技术优势，通过追查犯罪相关的数字轨迹，帮助被害人精准定位施暴者。同时，平台应对用户网络日志进行留存，对原始数据及时备份，尤其需要固定网络暴力的关键传播节点和舆情影响的关键信息，以便后续作为呈交行政执法或刑事追诉的证据。

其次，证据固定与协助取证义务成为平台治理中最关键的一环。《网络暴力信息治理规定》第二十五条明确要求，网络信息服务提供者应当及时保存信息内容、浏览评论转发数量等数据，并向用户提供网络暴力信息快捷取证等功能，依法依约为用户维权提供便利[25]。平台应利用技术优势，为受害者提供便捷的证据保全工具，如“一键取证”功能，帮助其收集侵权信息的发布时间、内容、传播路径、账号信息等关键证据，提高当事人证据收集的便利性，实现更加精确、有效的权利救济[9] (pp. 90-96)。在“杭州案”中，若谷女士能通过平台一键获取完整、规范的电子证据，其维权成本将大大降低。当司法机关介入后，平台有义务提供必要的技术支持和数据协助，查明网暴信息的源头及传播链条。

综上，构建平台“预警–控制–救济”的三阶段治理义务：平台事前的预防义务包括社区公约和“算

[8]“红旗规则”源自美国于 1998 年通过的美国版权法修正案，该修正案对平台的事后责任与触发的被动性予以强调，即平台等网络服务商在明确知晓网络侵权事实存在的情况下才需采取删除、屏蔽等系列措施以防侵权事实扩大化。该原则是对“避风港”原则一个重要的限制。

法对抗算法”技术治理，使得平台对网暴信息尽到合理注意义务，减轻其行政和刑事责任；事中平台主动采取控制措施包括限流、评论折叠、“紧急防护”功能等有效降低和阻断网暴信息扩散；事后的救济义务包括网上信息备份和协助取证义务，帮助受害者进行维权。平台只有进行信息生命周期的全流程监管，才能实现对网络暴力的实时、动态、有效治理。

5. 反向规制：平台刑事归责的具体模式与界限

如果说前述“事前、事中、事后”阶段化、体系化的内部治理义务是“正向激励”，那么，当平台急于履行或根本违反这些义务时，就需要刑法这根“高压线”进行“反向规制”，实现平台责任与刑事归责的真正贯通。但刑法介入应坚持其补充性和谦抑性原则，只有在平台的不作为严重侵害法益，且前置法已不足以规制时才应启动。

(一) 直接责任的激活

在网络暴力犯罪案件中，很多犯罪行为都是通过网络平台实施。例如，网络侮辱、诽谤案件往往都是利用网络平台进行群体性的、持续性诋毁、攻击、谩骂等。在网络暴力犯罪案件涉及的主体中，不仅包括发起者、组织者、参与者，还包括网络服务提供者，为此，网络暴力犯罪的惩治理应包括对网络服务提供者相关犯罪行为的治理，并且应强化对其刑事责任的追究。根据《刑法》第 286 条之一拒不履行信息网络安全管理义务罪的规定，如果网络服务提供者在经监管部门责令采取改正措施后拒不改正，那么对于所造成的违法信息大量传播等危害后果应当承担相应的不作为责任。此外《指导意见》第六条也明确指出，平台对于发现的网暴信息不依法履行安全管理义务，经监管部门责令改正而拒不改正，造成严重后果的，可以拒不履行信息网络安全管理义务罪定罪处罚。然而，该罪名因构成要件不明确、入罪门槛设置过高、行刑程序衔接不畅通等问题，在实践中一度被诟病为“僵尸条款”[30]。要激活这一条款，关键在于明确“信息网络安全管理义务”的具体内涵，并降低启动门槛，明确其程序衔接机制，以发挥该罪在惩治网络暴力犯罪案件中的应有功能。

此外，随着《网络暴力信息治理规定》的正式施行，前述的“主动预警”“动态干预”“协助取证”等义务被明确为平台的法定职责。平台违反这些义务，且在监管部门责令后，仍无正当理由拒不改正，就具备了构成此罪的可能。例如，某平台若已建立网暴识别模型，却为了流量收益故意关停或忽视其警报，放纵针对特定个人的诽谤信息大面积传播，最终导致被害人精神失常等严重后果，就完全可能落入本罪的规制范围。

(二) 间接责任的厘定

除前述对网络平台追究拒不履行信息网络安全管理义务的不作为责任外，还应加强对平台直接实施网络暴力犯罪实行行为和帮助行为刑事责任的追究。常见的情形是：平台没有明显违反监管部门的“责令”，但对平台内的网暴行为持放任甚至默许态度，例如平台出于蹭热度、推广引流的目的，放任甚至助推网络暴力犯罪行为的发展，此时就需要：追究其作为共犯，特别是“片面帮助犯”的刑事责任[31]。在网络暴力犯罪中，平台与具体的施暴者之间通常没有双向的意思联络，但平台明知，包括“知道”和“应当知道”用户利用其服务实施侮辱、诽谤、侵犯公民个人信息等犯罪，仍为其提供通讯传输、信息推广、技术支持等实质性帮助。此时，即使没有共谋，也可以认定平台构成“片面帮助犯”，按照其帮助的犯罪定罪处罚。例如，平台若明知某个账号长期、持续地发布针对特定个人的诽谤信息，且该信息屡次登上热搜，却因该账号是“大 V”、能带来流量而疏于管理，甚至反向助推，就完全可能构成诽谤罪的片面帮助犯。此外，若网络平台出于商业竞争等目的对其他平台或个人直接实施攻击、谩骂、诋毁等网络暴力犯罪行为，则应以相关罪名直接归责。

综上，应区分平台的不作为犯与实行犯、帮助犯，辨别拒不履行信息网络安全管理义务罪与其他帮

助型犯罪。对于网络服务提供者明知他人利用信息网络实施网络暴力犯罪行为，仍为其提供网络接入、信息推广等帮助行为的，可以帮助信息网络犯罪活动罪认定处理；对直接实施侮辱、诽谤、侵犯公民个人信息等网络暴力犯罪行为的，应以相关罪名进行规制。

(三) 责任的限缩

为防止因过度强调平台责任而侵蚀公民的言论自由，或给平台施加不切实际的过重负担，必须合理限缩数字平台的义务履行范围，为平台的刑事责任设定清晰界限，使其能够在合理的治理成本和风险管控范围内实现预期的治理效果。因此仅当平台违反法定的、明确的、技术上可行的作为义务，且主观上达到“明知”或“应知”的程度，方可启动刑事追责，避免不当干预言论自由和平台正当经营。

首先，应坚持刑法谦抑性原则。对于平台的不作为，若能通过行政处罚，如警告、罚款、暂停信息更新等能有效规制，或通过民事连带赔偿足以弥补受害人损失，就不应轻易启动刑事追诉程序。防止刑罚的“水波效应”可能波及平台的无辜员工和第三方合作商，抑制行业创新[32]。

其次，坚持主观“明知”的实质判断。平台的间接刑事责任以“明知”为前提[33]。对于用户的有效通知，应审查其是否符合法定形式要件。这里的“明知”包括知道和应当知道，而对于“应当知道”的判断，为防止责任泛化，应综合信息的热议程度、传播范围、是否位于醒目位置等因素进行实质判断，避免仅因海量信息存在就推定平台“应知”从而课以过重的审查义务[34]。具体而言，当出现以下情形时，可综合判断平台是否处于“应当知道”的状态。第一种情形看信息传播速度和规模。当网暴信息进入平台的热搜榜单、话题推荐等显著位置，或在 24 小时内传播层级超过 3 层，速度呈指数型增长、转发次数超 500 次以上时，出现该种异常，平台应具有高度的注意义务[8] (pp. 220-221)。第二种情形看内容的异常程度。当特定攻击性关键词在特定话题下密集出现，或同一账号被大量用户集中举报，平台的风控系统应当能识别此类异常模式，若此时平台未触发任何预警或处置，可推定其存在“应知”的过失[35]。第三种情形看平台的技术能力和行业标准。判断平台是否“应当知道”，还应结合同期行业采取的普遍技术审核标准。若大多数平台均能有效识别并处置类似信息，而涉案平台因其技术投入不足或管理漏洞未能发现，或未采取措施，则应认定其存在过失[36]。这实际将平台技术能力纳入“注意义务”范围。

最后，引入合规作为出罪事由[37]。如果平台已经建立了符合法律规范和行业标准的反网暴合规体系，并切实履行了事前预警、事中干预、事后救济等义务，即便仍有少量网暴信息未能被完全杜绝，也应认定其已尽到合理注意义务，阻却刑事责任的成立。为使合规出罪制度更具公信力与可操作性，可借鉴环境法、金融监管等领域的有益经验，引入由网信部门认可的、具备专业资质的第三方机构，定期对平台的网暴治理体系进行独立评估[8] (pp. 154-155)。评估内容应包括技术模型的有效性、执行落实的及时性、整改反馈的针对性等。平台可依据有效的合格评估报告，在发生少量网暴信息遗漏时主张“已尽合理注意义务”，作为刑事责任的抗辩事由。司法机关在审查平台刑事责任时，应将此评估报告作为判断平台主观上是否存在过错的重要参考。通过上述引入平台自治与第三方评估相结合，形成公私合作的治理闭环，既能有效遏制网暴风险，也能为平台提供稳定的行为预期。

当然，网络暴力治理不能单方面强调平台责任，基于风险管理的有限责任限缩体平台的义务履行范围，更能促进平台积极履责[38]。另外还需要结合政府、社会、个人等多元治理主体的协同参与，才能共同推动网络舆论生态整体改善，有效控制网络暴力发生[15] (pp. 129-130)。

6. 结语

网络暴力的治理，是一场技术与制度的赛跑。传统的、聚焦于事后惩罚单个施暴者的“点状打击”模式，已难以应对具有“涉众性”“累积性”和“平台依赖性”的新型网络暴力。本文通过对“杭州取快递被造谣案”的深度剖析，揭示了这一治理困境的根源，并论证了构建平台责任与刑事归责联动机制

的必要性与可行性。

这一机制的核心理念，是实现治理范式的根本性转移：从“国家–个人”的直接对立，转向“国家–平台–个人”的间接治理。即国家通过设定明确的法律义务和刑罚底线来“管平台”，而平台则利用其技术优势和“私权力”来“管用户”。在这个框架中，平台不再是中立的旁观者或消极的执行者，而是被塑造为积极主动、权责统一的“数字看门人”。通过“事前预警–事中干预–事后救济”的正向激励，将平台的技术能力转化为治理效能；通过拒不履行信息网络安全管理义务罪与片面帮助犯理论的“反向规制”，为平台不作为划定刑罚边界。这种双向互动，旨在将平台的商业利益与公共利益联系在一起，最终实现从数字平台网络暴力治理的角色转变。

当然，这一机制的设计必须慎重，以防止因过度强调平台责任而侵蚀公民的言论自由，或给平台施加过重负担。如何“管得住”又“放得活”相平衡，如何在不同类型的平台之间设定差异化的注意义务标准，如何在算法治理时代透过“技术黑箱”合理归责，将是未来理论研究与实践探索需要持续深化的重要课题。但无论如何，正视平台的力量并将其纳入法治轨道，已是数字时代网络暴力治理的必由之路[39]。可以说，让平台积极参与网络暴力治理已成为各界的共识，而本文所倡导的联动机制，正是迈向这一共识的制度桥梁。

参考文献

[1] 单勇. 论网络犯罪的看门人规制[J]. 南京大学学报(哲学·人文科学·社会科学), 2023, 60(5): 57-74, 157.

[2] 单勇, 李林. 网络暴力信息的看门人治理——以《网络暴力信息治理规定》为中心[J]. 学海, 2025(1): 90-100, 213-214.

[3] 单勇. 数字看门人与超大平台的犯罪治理[J]. 法律科学(西北政法大学学报), 2022, 40(2): 74-88.

[4] 王华伟. 网络暴力治理: 平台责任与守门人角色[J]. 交大法学, 2024(3): 51-64.

[5] 田宏杰. 网络暴力刑法治理中的“法不责众”困境及其化解[J]. 法学杂志, 2024, 45(1): 29-44.

[6] 蔡永成. 网络暴力的刑事治理逻辑及其路径[J]. 法治研究, 2025(5): 135-146.

[7] 蔡明函, 于冲. 涉众型网络暴力犯罪的刑事归责路径[J]. 江苏警官学院学报, 2025, 40(3): 40-49.

[8] 刘艳红. 网络暴力治理法治化研究[M]. 北京: 法律出版社, 2023: 86, 220-221.

[9] 张校基. 从个人到平台: 数字时代的网络暴力刑法规制[J]. 江西警察学院学报, 2025(1): 87-96.

[10] 于冲. 论网络聚合犯罪的刑法规制[J]. 中国法学, 2025(5): 164-183.

[11] 刘宪权, 周子简. “机器人”可否成为诈骗罪对象[N]. 检察日报, 2021-07-20(003).

[12] 刘宪权, 周子简. 网络暴力的刑法规制困境及其解决[J]. 法治研究, 2023(5): 16-27.

[13] 张云霄, 王泽南. 网络暴力刑事责任主体的界分与归责[J]. 中国检察官, 2025(19): 24-26.

[14] 魏琪. 网络暴力治理中的平台义务[J]. 荆楚法学, 2025(3): 50-63.

[15] 翟岩. 数字平台治理网络暴力的监管困境与角色重塑[J]. 华南理工大学学报(社会科学版), 2025, 27(4): 120-130.

[16] 翟岩, 李小波. 人工智能安全视域下“数据投毒”的内涵特征、层级风险与治理路径[J]. 河南社会科学, 2025, 33(11): 113-124.

[17] 翟岩, 李小波. 数字平台网络暴力: 内涵证立、生成机制与治理路径[J]. 科技与法律(中英文), 2025(3): 93-104.

[18] 冯明昱. 累积犯视阈下网络暴力的分层归责与治理机制研究[J]. 南昌大学学报(人文社会科学版), 2025, 56(5): 121-131.

[19] 龚文博. 论网络暴力治理中的平台间接刑事责任[J]. 法制与社会发展, 2024, 30(6): 147-162.

[20] 肖宸彰. 网络暴力刑事治理的归责逻辑重构与规范迭代[J]. 华东政法大学学报, 2025, 28(4): 178-192.

[21] 曾磊, 令宏亮. 网络暴力行为的平台规制困境与刑事应对[J]. 广西警察学院学报, 2025, 38(4): 70-79.

[22] 刘文杰. 网络暴力中个体维权困境与平台安全保障义务[J]. 国家检察官学院学报, 2023, 31(5): 39-57.

[23] 赵宏. 网暴案件中的民行刑责任与一体化衔接[J]. 北方法学, 2023, 17(5): 5-20.

[24] 石佳友. 网络暴力治理中的平台责任[J]. 法律科学(西北政法大学学报), 2023, 41(6): 14-23.
[25] 孔祥稳. 网络平台信息内容规制结构的公法反思[J]. 环球法律评论, 2020, 42(2): 133-148.
[26] 喻海松. 网络暴力的多维共治——以刑事法为侧重的展开[J]. 江汉论坛, 2023(5): 128-135.
[27] 赵镧淏, 刘晶. “规则 + 技术”: 互联网平台针对“网络暴力”现象的数字化治理[J]. 北京文化创意, 2023(2): 75-84.
[28] 王立梅. 网络空间下避风港原则的完善与网络服务提供者责任分类[J]. 江西社会科学, 2020, 40(5): 157-167, 256.
[29] 戴长林. 网络犯罪司法实务研究及相关司法解释理解与适用[M]. 北京: 人民法院出版, 2014: 100.
[30] 周立波. 网络暴力犯罪的刑事法治理[J]. 法治研究, 2023(5): 38-51.
[31] 龚文博. 网络平台处理个人信息的合规义务及其出罪路径[J]. 华东政法大学学报, 2024, 27(2): 52-66.
[32] 姜涛. 网络暴力治理中刑事责任、行政责任与民事责任的衔接[J]. 法律科学(西北政法大学学报), 2023, 41(5): 102-114.
[33] 敬力嘉. 网络服务提供者的间接刑事责任——兼论刑事责任与非刑事法律责任的衔接[J]. 网络法律评论, 2016, 20(2): 146-166.
[34] 刘金瑞. 网络暴力侵权法规制路径的完善[J]. 政法论坛, 2024, 42(3): 66-76.
[35] 赵健旭. 网络暴力识别预警机制的法治建构[N]. 民主与法制时报, 2022-05-11(003).
[36] 班天可. 安全保障义务的边界——以多伊奇教授对交往安全义务的类型论为视角[J]. 中德法学论坛, 2017(2): 168-183.
[37] 宋佳宁, 华琪. 网络暴力行为的刑法规制困境探析及应对[J]. 中国发展, 2025, 25(3): 26-33.
[38] 朱笑延. 从内容监管到生态调控: 网络暴力信息治理的平台义务重塑[J]. 南京大学学报(哲学·人文科学·社会科学), 2024, 61(1): 76-91, 163-164.
[39] 喻海松. 刑事一体化视野下网络暴力的规制模式[J]. 法律科学(西北政法大学学报), 2023, 41(5): 83-91.