

数字平台网络暴力的内生逻辑与规制进路

罗心语 韩紫阒 陆诗享 沈家仪 张桂莹 胡卓焯 程刚 童银辉^(通讯作者)

浙大宁波理工学院, 浙江省宁波市, 315100;

摘要: 随着数字平台日益成为社会信息交互的核心场域, 网络暴力的生成机制、传播模式与危害形态均已发生深刻变化。平台不再只是信息传输的中立渠道, 其技术架构、商业模式与运营规则深度渗透于网络暴力的全过程, 成为塑造这一生态的关键因素。在此过程中, 平台所扮演的多重角色与相应的责任张力亟待澄清。算法推荐放大争议内容、流量经济激励极端言论、以及平台在内容审核与用户保护方面的制度缺陷, 均暴露出传统“避风港”原则应对复杂网络暴力时的局限。因此, 网络暴力的治理须推动从“管道责任”向“看门人责任”的范式转型, 进而构建以法律义务为底线、以算法治理为核心、以多元协同为支撑的平台责任体系。

关键词: 网络暴力; 平台责任; 算法治理; 协同治理

DOI: 10.69979/3041-0673.26.03.094

1 平台化社会与网络暴力的发展历程

互联网的发展推动社会从分散的网络形态转向高度集中的平台模式。数字平台以其数据整合、中介协调与即时互动的特点, 重构了信息生产、传播与消费的链条, 形成新型数字公共领域。然而, 这一过程也催生了新的社会风险, 网络暴力便是其中典型。在平台社会中, 网络暴力已不再限于早期的辱骂或诽谤, 而是融合了技术操纵、群体动员、情感煽动与资本运作的复杂形态。它的发生和扩散高度依赖各大数字平台, 并与算法推荐、流量分配及社区互动机制紧密相连。

数字平台在此过程中呈现出双重角色。一作为基础设施提供者, 平台为用户提供表达与连接的可能; 二作为拥有市场和组织能力的私主体, 平台的技术设计与运营策略事实上构成了网络暴力的发生条件。例如, 为提升用户黏性与停留时长, 平台算法常优先推送高唤醒、易引发对立的内容。这类流量导向的机制, 客观上为攻击性言论和失真信息的传播提供了便利, 使其能够快速跨越圈层, 形成大规模舆论事件。

数字平台社会中的网络暴力实质上是技术与社会结构互动的结果, 其中平台的基础性权力成为理解其生成与传播的关键。这种权力不仅体现于信息分发, 也隐藏于界面设计、社区规则与激励体系之中, 潜移默化地影响着公众的交往逻辑与价值判断。因此, 要应对网络暴力在平台时代的蔓延, 不仅要依靠法律与伦理约束, 更需推动平台突破流量主导的运行逻辑, 转向具有公共责任感的社会化治理。

2 平台在网络暴力生成与扩散中的结构性作用

网络暴力的平台化演进, 意味着对其根源与动力机制的研究必须切入到平台经济的运行逻辑与技术特征这一核心视角。平台并非被动容纳暴力的容器, 而是通过一系列主动或被动的选择, 在一定程度上参与了网络暴力的塑造。

首先, 算法推荐系统构成了网络暴力信息分发的核心引擎。基于协同过滤、用户画像和实时热度的推荐模型, 其首要目标是提升用户参与度与平台总体流量。在这一目标驱动下, 能够引发强烈情绪反应(如愤怒、鄙视、对立等)的内容更容易获得算法的青睐, 被赋予更高的权重和更广阔的曝光机会。这导致一个悖论: 即使平台在内容政策中明确反对暴力, 其底层的技术逻辑却在事实上奖励暴力。当关于某个人或事件的分裂性观点出现时, 算法会迅速识别其“争议价值”, 并将对立双方的内容分别推送给立场相近的用户群体, 加剧信息茧房效应和群体极化。用户被困在由算法编织的、不断自我强化的同质化信息环境中, 理性对话的空间被压缩, 对“异己”观点的容忍度下降, 为跨群体的言语攻击埋下伏笔。算法在其中的作用, 从单纯的信息匹配工具, 异化为塑造舆论氛围、设置公共议程的隐性权力。

其次, 平台的商业模式与社区治理规则直接影响了施暴的成本与收益。在流量变现的激励下, 部分意见领袖(KOL)、营销号甚至普通用户发现, 发表极端、片面或煽动性言论往往是获取关注、积累粉丝的捷径。平台虽然建立了举报和审核机制, 但在实践层面常面临效率与精准度的挑战。对于大量游走在法律边缘、采用“阴

阳怪气”、隐喻反讽等软暴力手法的内容，自动化审核模型识别困难，而人工审核又难以应对海量级的内容生产。这种审核滞后与违规成本低廉的现状，变相降低了网络施暴的心理门槛与实际风险。施暴者可能仅需承受账号被短暂限制或内容被删除的轻微后果，而他们制造的舆论海啸却已对受害者造成不可逆的精神伤害与社会关系破裂。平台在治理资源的分配上，往往更侧重于事后处置而非事前预防与事中干预，这种“重响应、轻防控”的模式难以应对网络暴力瞬时爆发、链式扩散的特点。

最后，平台提供的技术工具与交互功能，为新型暴力形式的出现提供了可能。从早期的“人肉搜索”到如今的“开盒”（利用技术手段非法获取并曝光他人个人信息），从文字辱骂到恶意剪辑的短视频、伪造的音频，技术手段的迭代不断翻新着网络暴力的样态。平台作为这些技术工具的提供者与应用场景的搭建者，虽然难以预见每一项具体的技术滥用，但负有通过产品设计、权限管理、风险监测等手段尽可能降低技术被用于作恶可能性的义务。然而，在用户增长与市场竞争的压力下，平台在便捷性与安全性之间的天平有时会向前者倾斜，例如对某些数据接口管理不严，对内容合成工具缺乏有效溯源标识等，这在客观上为技术化、组织化的网络暴力提供了便利。

3 现行法律框架下平台责任的困境与局限

面对平台在网络暴力中的结构性作用，我国现行的法律规范体系在界定与追究平台责任时，显露出明显的滞后性与适配困境。当前规制模式仍深受传统“避风港”原则的影响，其核心思路是将平台视为中立的技术服务提供者，责任触发通常以“通知-删除”为前提，即平台在接到权利人明确通知后，若及时移除侵权内容则可免责。这套机制在 Web1.0 时代对于处理明确的版权侵权等问题曾发挥一定作用，但在应对复杂、动态、群体性的网络暴力时则捉襟见肘。

困境首先体现在归责逻辑的错位。“避风港”原则隐含的是一种事后救济、被动响应的责任逻辑。然而，网络暴力的危害往往不在于单一侵权信息的持续存在，而在于信息经由平台推荐系统引发的爆炸式、不可控的传播过程及其累积的社会效应。受害者可能因一段被恶意曲解的视频在几小时内遭受数万条攻击性评论，即使事后平台删除了原视频，其造成的心理创伤与社会性死

亡后果已难以挽回。要求受害者逐一识别并通知海量的侵权信息发布者及传播链接，在实践中不具备可操作性。平台凭借其技术能力，本应对其系统内信息的异常传播态势具有更高的预见和管控可能，但现行法律并未对这种基于技术优势的积极防控义务作出充分要求。

其次，责任标准模糊导致司法与实践之间的摇摆。我国《民法典》《网络安全法》《个人信息保护法》等法律虽原则性地规定了网络服务提供者需履行安全管理义务，但何种情况下平台应为其系统内传播的暴力信息承担责任，特别是当其并未直接创作内容，而是通过算法进行了推送时，判断标准并不清晰。《关于依法惩治网络暴力违法犯罪的指导意见》等文件试图作出细化，但如何认定平台对“明知或应知”的暴力信息未采取必要措施，仍存在很大解释空间。平台往往以其采用自动化技术、无法实时审查所有内容为由进行抗辩。这导致在司法实践中，平台责任认定的结果具有较大不确定性，难以形成稳定的预期和有效的威慑。同时，行政处罚力度与平台从流量经济中获得的巨大收益相比，有时显得威慑不足，难以驱动平台投入足够资源进行根本性的治理模式变革。

更深层的矛盾在于，现有责任框架未能有效回应算法时代平台角色的根本转变。当平台的算法深度介入信息的分发与排序，甚至无形中塑造了用户的认知偏好与舆论走向时，继续简单套用针对“通道”或“存储空间”的旧有规则，显然已不合时宜。算法并非完全中立的技术工具，其设计蕴含着开发者的价值选择与商业目标。当算法为了追求“用户 engagement”而系统性偏好具有冲突性的内容时，平台能否以“技术中立”为由完全免责，已成为法学与伦理学争论的焦点。本文认为，平台对其核心算法带来的系统性风险，应承担起相应的治理责任，这要求法律责任的设定必须穿透技术的“黑箱”，触及算法设计与部署的伦理与合规层面。

4 以平台为核心的责任重构与协同治理体系

由于网络暴力日益依赖平台传播而现有规制手段效果有限，治理思路必须进行调整。平台的责任定位应当从主要进行事后内容删除，转向更积极地预防风险与参与生态建设。要实现这一目标，需从法律规则更新、平台运营调整以及多元主体合作等方面系统推进。

现行法律体系中的“避风港”原则已难以应对平台社会中的复杂挑战。因此，应确立与其技术能力和市场

地位相称的“看门人”义务，尤其是对那些具有显著公共影响力的大型数字平台。这意味着立法需设定更高的注意义务标准。平台应建立与其服务规模匹配的高效审核机制，提升对煽动对立、隐性贬损等复杂内容的辨识水平。算法设计需要避免过度放大争议与仇恨，平台有责任建立算法影响评估与社会风险报告的常规机制。同时，必须完善用户权益保护渠道，为受害者提供便捷的投诉、取证与支持路径，并明确规定平台响应投诉的时限与流程。对于因未尽到合理注意义务而导致网络暴力广泛传播并造成严重后果的，法律应规定具有足够威慑力的行政处罚，并探索通过民事公益诉讼等路径进行追责。此外，可以鼓励平台建立用户行为记录机制，对反复参与施暴的账号采取阶梯式限制措施。

平台自身需要将社会治理目标内化到商业运营和产品设计之中。提升透明度是关键，平台应在不泄露核心技术秘密的前提下，向社会说明其推荐逻辑、流量分配原则及内容处置标准，接受外部监督。资源应投入于开发更精准的风险预警模型，实现对潜在暴力舆情苗头的早期识别与干预，将治理节点前移。审核团队需要持续的专业培训，以提升其在复杂语境下的判断力，并应获得必要的职业保障。更为根本的是，平台必须审视并调整其商业模式的内在激励，改变单纯以流量和互动数据为核心的评价体系，将内容质量和社会责任等多元指标纳入考量，从根源上降低对极端言论的经济激励。

网络暴力的治理最终依赖于政府、平台、行业组织、媒体、学校及公众共同构建的协作生态。政府监管部门应加强跨部门协调，并与主要平台建立数据共享与执法联动机制，以提升监管实效。行业组织可以牵头制定更细致的自律准则，并推动建立跨平台的违规联动处理与从业者信用记录。媒体在报道相关事件时必须恪守专业伦理，避免渲染对立，引导公众进行理性讨论。教育体系应将网络素养与法治教育纳入课程，帮助学生形成负责任的数字行为习惯，相关培训也应覆盖教师与家长群体。网民个体需要提升信息辨别能力，在参与讨论时保持审慎，主动抵制暴力信息并善用举报渠道。作为关键节点的平台，应定期向社会公开治理情况，并将公众评议反馈到运营之中，努力在商业目标与公共利益间寻求

平衡。唯有通过这种多主体、持续性的协同努力，方能构建起遏制网络暴力的长效机制，推动数字空间向更为理性、包容的方向发展。

5 结语

平台化时代网络暴力的蔓延，实质上是技术条件、商业逻辑与公共规范相互作用的结果。作为关键节点的数字平台，其责任范畴已不能局限于“通知—删除”这一传统框架，而必须正视平台通过算法架构、商业模式与规则设计对网络暴力生态的形塑作用，并据此重塑其法律责任与社会角色。在未来的治理中，应坚定推动平台责任范式的现代化转型，以强有力的法律规范明确其“看门人”角色，以持续的伦理审视约束其算法权力，以开放的协同共治激发多元主体的参与活力。这一治理转型旨在数字时代平衡表达自由与人格尊严的保护。只有通过系统性的责任重构与治理升级，才能有效应对平台时代的技术风险，在保障表达自由的同时，切实维护个体在网络空间中的尊严与安全，共同营造清朗有序的数字环境。这不仅是应对当前网络暴力挑战的必然路径，也是构建更具包容性、安全性和公正性的数字公共空间的坚实基础。

参考文献

- [1] 董玮, 吴欣宜. 人工智能时代互联网平台算法责任研究——基于短视频平台网络暴力的法律规制[J/OL]. 新媒体与社会, 2025.
- [2] 姜涛. 网络谣言的刑法治理：从宪法的视角[J]. 中国法学, 2021(3).
- [3] 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(03).
- [4] 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).
- [5] 余梦君. 自媒体衍生网络暴力问题与治理路径[J]. 新闻世界, 2025, (06).
- [6] 魏琪. 网络暴力治理中的平台义务[J]. 荆楚法学, 2025, (03).
- [7] 邝岩. 社交媒体平台网络暴力防治效果评价指标构建及实证研究[J]. 情报杂志, 2025(44).