武漢大學學報(哲学社会科学版)
JOURNAL OF WUHAN UNIVERSITY(Philosophy and Social Sciences)
第 79 卷第 5 期　2026 年 9 月
Vol. 79 No. 5　Sept.　2026　164~176
DOI:10.14086/j.cnki.wujss.2026.05.015

论网络暴力的生成机制与刑法规制

马永强

摘　要　深入探究网络暴力的生成机制,是有效规制该现象的逻辑起点与现实基础。究其根源,网络暴力在生成机制上受社交网络平台结构性缺陷、现实社会冲突的网络化呈现以及网络暴力与现实身份的关联性等多重因素影响。反思既有治理进路,技术治理模式虽不可或缺但非治本之策,其治理效能及潜在风险亟待审慎评估,而夯实法律治理模式是解决问题的关键。为此,应基于网络暴力的生成机制,构建以领域立法为基石、跨领域协同为保障的综合治理体系,明晰刑法介入限度并审慎对待新罪增设。在刑法规制进路的优化上,应立足于网络空间治理的特殊性,通过刑法适用强化网络平台的刑事责任;需充分探究关键信息传播者及其他网络暴力参与者的责任约束;要反思累积犯适用路径,恪守罪刑法定原则,通过对既有罪名的扩张解释与必要的立法完善,切实提升刑法规制能力。

关键词　网络暴力;平台责任;领域法;信息网络安全管理义务;网络开盒;刑法解释
中图分类号　D913　**文献标识码**　A　**文章编号** 1672-7320(2026)05-0164-13
基金项目　国家社会科学基金青年项目(23CFX041)

随着社交平台深度嵌入社会生活,网络暴力已由零散的言语失范异化为互联网治理领域的顽疾,并呈现出高发化与常态化的态势。其危害亦已从对个人生活安宁的滋扰,演变为可能诱发受害者精神崩溃乃至自杀等严重后果,因此,治理网络暴力的紧迫性持续凸显。然而,既有刑法学研究多着眼于法教义学的规范分析,对网络暴力赖以滋生的深层生成机制探讨有限,致使规制方案与网络空间的现实运作机理之间有所脱节。有鉴于此,亟须引入交叉学科视野,立足网络暴力的复杂生成机制,在审慎反思技术与法律治理既有进路的基础上,厘清刑法的功能定位与介入限度,为网络暴力的刑法规制提供支撑。

一、问题的提出

近年来,因网络暴力导致受害者深陷舆论旋涡、最终绝望轻生的悲剧性事件屡见不鲜,持续引发舆论关注,网络暴力的现实危害进入公共视野。究其本质,网络暴力并非单纯的线上谩骂,而是施暴者借助网络平台的匿名性、便捷性和放大效应,对个体实施蓄意、持续且集中的攻击,进而造成现实损害的新型网络侵害行为。依据国家互联网信息办公室等四部门出台的《网络暴力信息治理规定》(以下简称《治理规定》),网络暴力涵盖了侮辱谩骂、造谣诽谤、煽动仇恨、威逼胁迫、侵犯隐私,以及影响身心健康的指责嘲讽、贬低歧视等多种行为样态,可见网络暴力在行为方式上的多样性;其行为动机亦可划分为个体攻击型、报复社会型与解构权威型等类型[1](P84)。

网络暴力虽然发生在虚拟空间,侵害对象却指向现实世界的个体及其切身权益,具有现实性和群体性,其后果亦具有破坏性。网络暴力系网络空间中的群体性事件,具有高度的传播性、煽动性和道德指向性。施暴者常利用网络的匿名性和扩散性对受害者进行围攻和污名化,形成“多数人的暴政”。其不仅严重侵扰个人生活安宁,造成人格贬损、声誉破坏、社会排斥等严重后果,甚至会诱发受害者自杀,侵

蚀社会信任、公共秩序与网络生态，滋生仇恨情绪并加剧社会撕裂。

针对网络暴力乱象，最高人民法院、最高人民检察院、公安部联合发布了《关于依法惩治网络暴力违法犯罪的指导意见》（法发[2023]14号，以下简称《指导意见》），为刑法规制提供了规范依据。尽管如此，对于网络暴力何以滋生蔓延、刑法应如何介入以及如何准确划定刑法介入限度等议题，学界迄今尚未形成共识。因此，须深入剖析网络暴力的复杂生成机制，以廓清上述争议并提升刑法对现实治理需求的回应能力。

二、网络暴力的复杂生成机制

网络暴力并非单纯的网络失范，其生成机制根植于网络空间与现实社会的复杂互动。将其简单归咎于个体道德缺失或网络平台监管不足，或套用犯罪学理论进行表象描述，均无助于揭示其本质。应当深入探究网络空间的技术逻辑、社会结构与群体心理，以廓清网络暴力的风险生成与扩散机理。社交平台的结构性缺陷、现实冲突的网络投射以及网络暴力对现实身份的攻击，共同构成了网络暴力生成机制的核心要素。

（一）虚拟场域：网络暴力生成的结构性诱因

网络暴力的生成与社交网络平台这一虚拟场域的结构性特征紧密相关。平台作为一种新型数字权力的集中体现，正通过无形的引导与规训，将个体塑造为温顺且可预测的数字公民[2]（P160）。Web 2.0时代，网络平台崛起使互联网的信息生产模式由精英化转向大众化，前端交互技术演进亦降低了信息发布门槛，使用户得以主导内容生产，这为网络暴力的滋生提供了技术与社会土壤。加之移动互联网与视频平台的普及，用户规模急剧膨胀，网络言论质量参差不齐，为网络暴力滋生提供了多样诱因。同时，社交网络平台作为连接施暴者和受害者的关键媒介，本应承担“守门人”的职责，却因商业逻辑与技术特性的裹挟，沦为网络暴力蔓延的温床。平台通过对信息流向和用户行为的隐蔽操控，形成了一种基于技术和资本的新型支配性权力。其权力偏好与运作机制，深刻影响着网络暴力的生成。

首先，平台算法的不当激励与议程设置的结构性异化，协同驱动了网络暴力的滋生及蔓延。为攫取流量红利与商业收益，平台推荐算法往往偏向于分发耸人听闻、让人情绪化的内容，这实质上构成了对煽动性、攻击性内容的隐性激励[3]（P73-74）。部分平台参与者甚至与网络黑灰产从业者相互勾连，雇用专业的黑公关团队批量注册僵尸账号、制造虚假互动，利用平台算法漏洞实施恶意营销炒作或网络攻击，散布煽动性信息、引导舆论风向，人为炮制或放大网络暴力事件。社交网络平台议程设置的结构性异化，亦助长了网络暴力蔓延。议程设置作为一种技术权力运用形式，是指平台通过算法推荐、流量倾斜、内容优先级排序等技术手段，影响用户对信息的接触顺序、频率和重要性的判断，进而塑造用户认知、引导舆论走向。其在客观上助长了群体对立、情绪极化与信息操纵。例如，热搜榜单和话题标签等机制深刻影响事件的曝光度和传播路径，塑造着公众对网络暴力的认知框架和情绪反应。在此机制下，特定利益群体亦可利用议程设置操纵信息传播方向，加剧特定群体遭受网络暴力的风险。不同利益集团在博弈中使得议程设置竞争白热化，甚至可能借机制造替罪羊操控社会情绪，推动网络暴力事件扩散。

其次，平台传播结构的扁平化及其伴生的信息茧房与群体极化效应，显著增强了网络暴力的组织性和情绪动员能力。一方面，扁平化的传播机制将用户包裹于同质化信息流中，形成“信息茧房”，孕育出带有排异性和攻击性特征的亚文化，为网络暴力创造出潜在土壤。在茧房内部，用户倾向于选择性接触符合自身偏好的信息并屏蔽不同见解。这种群体同质性加剧了认知偏差与群体极化，甚至会催生极端主义、仇恨与暴力[4]（P57），致使施暴行为在特定社群内被正当化。另一方面，信息爆炸带来的“竞争性真相”导致信息的权威性与真实性被消解。各类真假难辨的“误导性真相”凭借其情感感染力或对特定用户群体心理的迎合，在叙事框架与价值认同的博弈中获得病毒式传播[5]（P324-325），助长攻击性言论。加之个体在信息素养和情绪驾驭能力上的不对等，理性对话空间被进一步挤压，个体宣泄极易演变为集体施暴。攻击者常利用社会热点，裹挟公众情绪对受害者进行道德审判，以群体之名将个体推入深渊。

最后,网络传播的幂律特征、匿名性和便捷性协同放大了网络暴力的破坏力。网络空间的影响力遵循幂律分布规律[6](P125,144),以关键意见领袖(以下简称"KOL")和网络信息内容多渠道分发服务机构(以下简称"MCN机构")为代表的少数枢纽节点,依托其粉丝基础可瞬间实现信息的广域覆盖,急剧放大网络暴力的声势与规模。同时,网络传播的匿名性降低了施暴成本,使个体更易摆脱现实社会身份和道德约束,肆意宣泄负面情绪,形成"去抑制效应"。在此效应下,个体责任感被削弱,攻击性行为普遍化。攻击者将受害者去人格化,视其为抽象的"异类"而非具体的人,以减少同理心与负罪感[7](P35-36)。网络传播的便捷性则使攻击性言论得以迅速传播至更广泛的受众中,使得对受害者的伤害指数级增加。

(二)现实映射:现实社会冲突的网络化呈现

网络暴力的生成根植于现实中的社会冲突,并在网络空间中被重塑及放大。网络空间的舞台特性加剧了个体的身份认同焦虑,使其情绪表达更趋极端化。价值观冲突、利益纠纷等社会冲突投射于作为现实社会结构镜像的网络空间,并借助群体互动机制加速扩散,使网络暴力行为被强化。这些现象不仅是网络互动的副产品,更是社会压力、群体心理与权力结构在虚拟空间的延伸。

首先,社交网络平台的舞台特性强化了用户的身份认同焦虑与归属感焦虑,成为网络暴力滋生的重要诱因。在此类展示自我和寻求认同的舞台上,用户精心修饰在线形象,编织叙事以迎合受众,甚至暴露他人隐私以获取点赞和关注。对外界认可的过度依赖加剧了个体的心理脆弱性,如身份的不确定性引起焦虑不安,网络互动也因表演行为而流于表面,缺乏深度和厚度[8](P290-291)。社交网络平台的点赞、评论、粉丝数量等量化反馈机制将这种"表演"属性结构化,持续激励个体为迎合受众而编织"剧本"[9](P139-140)[10](P214-215)。然而,对认可的持续追求反而加剧了身份认同的不确定性:当理想化的"人设"与现实自我或受众期待产生落差时,则极易招致攻击和谩骂,进而演化为网络暴力。社交网络平台的舞台特性亦加剧了归属感焦虑。为融入特定网络社群,个体不仅被动迎合主流话语,甚至主动攻击"异见者"来确证群体身份,如"饭圈"群体为维护偶像形象而主动"引战"。这极易激发并放大个体内心深处潜藏的偏见、歧视与仇恨等负面人性因素,并将之投射于网络言行,成为网络暴力的诱因。

其次,网络暴力作为现实社会结构性矛盾的网络镜像,其滋生与蔓延也源于现实冲突的线上投射。现实世界中的社会组织形态、权力结构、社会心态等都会被带入网络空间并借助社交网络平台加速发酵。越轨理论指出,社会结构压力与个体机会的背离易诱发越轨行为,网络暴力正是这种社会失范在网络空间的体现。现实社会中资源、地位等的不平等在网络空间中被激化,可能使网络暴力异化为压制弱势群体、巩固既有权力结构与话语体系的工具。网络空间亦成为现实情绪的"泄压阀",部分在现实中被边缘化的个体,转而在网络上攻击他人以宣泄情绪、寻求关注,获取虚假的权力感。社会竞争加剧与个体心理韧性减弱,更使得网络暴力的侵害对象呈现出向弱势群体和普通人扩散的弥散化趋势。因此,不能脱离社会结构与网络空间治理结构来理解网络暴力,必须深究其社会根源与权力运作机制。

最后,网络空间的群体互动机制加剧了责任稀释,是网络暴力迅速蔓延的另一重要因素。参与主体的多样性使网络暴力呈现出波及范围难以控制的群体性特征,既加剧了信息茧房效应,也使得一些现实中的非主流观点在网络空间被不成比例地放大声量,营造出虚假的共识,甚至诱发网络民粹主义,加剧群体对立。同时,网络暴力的群体性特征促成了责任稀释,其机制可借由旁观者效应和沉默螺旋理论予以解释。前者表明,旁观者数量增加反而导致个体因期待他人干预而陷入集体不作为的困境。后者则揭示,当个体察觉反对网络暴力的声量居于弱势时,会因担心被孤立或深感个体行为无足轻重而选择沉默。这使得施暴言论在互动中逐步占据主导地位,而理性反对网络暴力的声音最终被边缘化。

(三)身份关联:网络暴力破坏力的线下延伸

网络暴力虽发轫于虚拟世界,但其生成、演进及风险弥散却与现实社会结构紧密相连,呈现出虚实交织的特征。理解其破坏力的关键,在于充分认识其与受害者现实身份的关联性。网络暴力并非仅止步于虚拟性的语言霸凌,还可能通过现实身份的关联机制将伤害延伸至现实生活并造成严重后果。

首先，网络暴力行为与受害者现实身份的关联性体现为网络暴力对受害者现实生活的直接冲击。网络暴力的破坏力不仅源于施暴者借助虚拟身份的掩护，规避现实责任，实施肆意攻击，更在于其能够突破虚拟与现实的界限，将攻击指向受害者的现实身份。受害者的现实身份信息（如姓名、职业、住址、社会关系等）一旦被泄露，网络暴力便会直接波及受害者的现实生活，甚至可能使受害者遭到线下攻击，直接危及人身安全。网络的放大效应使得网络暴力信息迅速传播，难以被控制和消除，如同持久的标签，对受害者的现实生活造成持续性影响。

其次，这种关联性还会导致受害者经历“身份撕裂”和“社会性死亡”。持续的网络攻击及其现实影响会使受害者在其网络负面形象与真实自我之间产生剧烈冲突，导致其陷入身份认同危机，自我认知扭曲，最终走向“身份撕裂”。同时，作为社会生存基石的名誉和社会评价一旦崩塌，受害者在现实生活中亦会被孤立、排斥，甚至其社会根基和人际关系遭到毁灭性打击，陷入“社会性死亡”。一些受害者因网络暴力患上抑郁症、焦虑症等心理疾病，甚至自杀，这正是“身份撕裂”和“社会性死亡”的极端后果。

最后，个人信息保护的缺失和网络黑灰产业的猖獗使得这种关联性愈发增强，后果更加严重。互联网时代的过度分享文化无形中增加了个体隐私暴露的风险敞口，网络黑灰产业的猖獗更是令大量敏感个人信息被海量非法获取和泄露，受害者的个人隐私一旦被公开，便极易成为网络暴力的靶子。

三、网络暴力治理既有进路的问题及反思

网络暴力的复杂生成机制决定了其治理的艰巨性，对既有治理进路提出挑战。技术治理与法律治理各有优劣，有效的治理策略应在洞悉网络暴力生成机制的基础上，探索技术与法律并重的路径，并通过专门立法与跨领域协同夯实法律治理的根基。在刑法规制层面，则须厘清刑法介入限度、审慎评估新罪增设的必要性，确保刑法介入与网络暴力行为的法益侵害程度相适应。

（一）从技术治理模式的局限性看网络暴力治理的问题

面对传统法律手段在网络暴力治理中的应对乏力，技术治理模式备受推崇。当下，监管部门和网络平台正在探索“大V实名制”，并引入信息内容与高危账号识别模型等技术治理手段，《治理规定》亦强调运用人工智能、大数据等技术手段赋能信息识别与风险预警。技术治理在响应速度、处理规模等方面具有显著优势，但其本身固有的局限性与异化风险亦衍生出如下新问题：

首先，前台实名制可能加剧隐私泄露风险，并对网络言论生态产生复杂影响。辨析网络实名制对网络暴力的影响，需区分后台实名和前台实名两种不同模式。我国现行监管法规确立的后台实名制，要求用户向平台提供真实身份信息进行认证，但该信息不对外公开，《治理规定》第八条即属此列。为强化言论责任、降低虚假信息和网络暴力的风险，前台实名制亦进入公众视野。2023年10月31日，多平台推出“大V实名制”，要求拥有百万以上粉丝的自媒体账号，尤其在时政、财经等专业领域，率先公开真实身份[11]。后台实名制有助于平台追溯网络暴力源头、威慑潜在施暴者，对维护网络秩序确有裨益，但针对“大V”等特定群体的实名制，其治理效能与负外部性却亟待审视。前台实名制将用户的线上线下身份强行绑定，加剧了个人信息泄露及滥用的风险，极易导致网络暴力从匿名的言语攻击升级为更精准的现实生活骚扰，这显著增加了“大V”等公众人物遭受恶意骚扰乃至线下暴力等人身侵害的风险。

如前所述，网络暴力根源复杂多样，并非简单的匿名性所致，故而实名制难以触及根本。即使初期具有震慑作用，长期效果也难以维系，网络暴力可能变换形式继续存在，甚至变本加厉[12]（P157-158）。强制消除网络空间的匿名性不仅难以根治网络暴力，反而可能损害网络生态的健康发展。前台的匿名性作为网络空间的重要赋权特征，对保护隐私、鼓励个体在无后顾之忧的情况下参与公共讨论至关重要。强制推行前台实名制可能引发自我审查，形成“寒蝉效应”，最终侵蚀多元包容的网络舆论环境[13]（P34）。因此，技术治理手段的选择应审慎权衡不同利益，不宜以过度牺牲公民基本权利为代价。

其次，信息内容识别模型难以精准界定“不良信息”，易导致误伤正常言论并压制合理批评，进而减

损公民的言论自由。尽管《治理规定》将影响身心健康的指责嘲讽、贬低歧视等内容的违法和不良信息纳入广义的网络暴力信息范畴,但如何界定“不良信息”本身即是难题。尤其以冷嘲热讽、煽风点火、风言风语、阴阳怪气等表现形式呈现的不良信息型网络暴力,更是难以界定与治理[14](P97)。技术上,信息内容识别模型虽依赖自然语言处理等技术,但网络语言的高度复杂性、快速变异性和语境依赖性,使其难以捕捉微妙的语境差异,容易导致误判。

部分社会议题包含激烈批评,但在特定语境下仍属合理的公民表达,却可能被模型误判;某些看似“客观理性”的言论在特定情境下亦可能构成实质性伤害。如在“刘某某自杀案”中,一些网民对其寻亲行为和家庭背景进行看似“理性”的揣测和评论。这些“温和”的言论因缺乏同理心和对个体遭遇的尊重,客观上对当事人造成巨大的心理创伤,但无法被内容识别模型准确识别。即使人工审核介入,审核人员的误判、审核的形式化与保守立场,同样可能导致对正常言论的错误裁定。中文网络语境中普遍存在的缩写、谐音、隐喻、表情包等非标准化、高度情境化的表达方式,更增加了技术识别难度。如果过度屏蔽或错误屏蔽,不仅会干扰正常交流,也会在无形中压缩公共讨论空间,削弱公众话语权。因此,过度依赖技术手段识别不良信息,可能导致治理行为偏离比例原则,以规制之名不当干预甚至扭曲正常的网络舆论生态。

再次,账号识别模型的滥用可能对用户隐私保护构成严峻挑战,并加剧算法歧视与不透明性。此类模型的建构依赖对海量用户数据的深度分析,既涉及信息内容与历史行为,也涉及社交关系、地理位置乃至设备信息,因而必然伴随对用户隐私的收集和利用。其信息收集与利用的范围和深度,往往超出用户的预期和授权,甚至可能超出合理范围,构成对公民隐私权的侵犯。技术治理倡导者虽强调技术正当程序和比例原则[15](P132-133),但实践中却往往缺乏明确的界限和可执行的标准。在监管缺位与平台权力不受制约的语境下,技术逻辑将优先考虑风险识别效率,而忽视信息收集的必要性与合法性界限。受商业利益驱动,平台过度收集用户信息会增加隐私泄露风险,并侵害个人信息自决权。

账号识别模型并不中立,从设计、训练到应用等各环节均可能潜藏偏见。某些用户群体可能因其身份、地域、职业等固有属性或社会标签,更易被模型标记为高风险用户。而针对特定用户的系统性歧视,或对特定观点的不公正压制,则可能最终演变为隐蔽的“算法暴力”。尤须警惕的是,模型的黑箱化运作既阻碍了用户理解其决策逻辑,也削弱了外部监管,使用户难以有效申辩或寻求救济。实践中一些平台对算法运作机制讳莫如深,处置用户时又往往以“违反社区规定”等模糊说辞搪塞,不仅侵害了用户的知情权,更损害了技术治理的公正性与公信力。

最后,作为新兴技术治理工具的社交机器人与大数据技术,亦潜藏着异化与失控风险。基于生成式人工智能的社交机器人虽然被视为缓解网络暴力的潜在工具[15](P131),可用于辟谣或引导理性讨论,然而,如何确保信息生成的真实性与客观性,仍存在诸多挑战。一旦被平台滥用或被恶意操纵,社交机器人可能加剧虚假信息传播与舆论对立,甚至误导公共决策并加剧群体极化[16](P435-436)。大数据技术的应用同样存在数据偏见、隐私侵犯和安全隐患。倘若训练数据来源单一或未能充分清理歧视性内容,据此构建的模型便可能固化甚至强化既有社会偏见。平台过度依赖大数据技术进行用户画像、舆情监控乃至快捷取证,也容易诱发对用户数据的过度收集和滥用,违背数据处理的合法性、正当性与必要性原则。而大规模的数据泄露事件,则为精准诈骗、网络暴力等违法犯罪活动提供可乘之机。归根到底,现有技术治理路径的工具理性特质过于突出,使其在应对复杂人性、微妙语境及权利冲突时暴露出明显局限。其不仅未能给受害者提供全面有效的保护,甚至可能异化为新的治理风险源。倘若技术权力监管与伦理规范尚不健全,过度依赖技术治理和平台自律,将加剧网络暴力治理的复杂性。

(二)亟待以领域立法和跨领域协同夯实法律治理模式

技术治理模式既易被异化为平台牟利的工具,又难以触及网络暴力的深层社会根源与责任归属等根本问题,因此,应探索构建强有力的法律治理模式。在此进路之下,应充分考虑网络暴力的生成机制,形成以领域立法为基石、跨领域协同为保障的综合治理体系。

当前，我国的网络暴力治理尚缺乏系统性与前瞻性的法律框架。网络暴力成因复杂，其跨领域特征显著，难以归入单一部门法调整，制定专门领域法具有现实必要性。传统分散式立法模式的部门局限性较为突出[17](P5)，难以有效遏制网络暴力的蔓延，专门立法则能够清晰界定网络暴力行为的类型与程度，避免不同部门法之间的规范冲突和监管空白，进而提高治理的精准度。专门立法还有助于贯通事前预防、事中干预与事后追责，实现全链条治理。以领域法治思维为指引，制定治理网络暴力的专门性立法，有助于推动相关立法体系构建，全面实现网络暴力治理的法治化转型[18](P80)。专门立法亦有助于强化并压实平台责任。前已论及，网络平台的结构性缺陷是网络暴力滋生的重要根源，因此，明确平台的法律责任与监管义务，成为夯实法律治理模式的关键环节。在算法设计、内容推荐、信息审核、用户管理等诸多方面，专门立法可细化平台义务与责任约束，引导平台建立预警及处置机制，并提供维权途径。

除专门的领域立法外，完善跨领域协同，强化算法规制、保障个体权益、划定治理底线，是网络暴力法律治理模式构建的另一重要环节。领域法与部门法并非替代关系，而应是互补与协同的共赢关系[19](P7)。因此，有效规制网络暴力，须在完善领域立法的同时，优化现有法律规范的解释与适用。具言之，应基于对网络暴力的类型甄别，以民法、行政法、刑法齐头并进的方式进行跨领域的综合治理[20](P299)，细化民事责任、行政责任和刑事责任的衔接机制，形成有助于有效落实法律责任的多层次治理体系。基于网络暴力的生成机制，跨领域协同机制的构建应聚焦于以下方面：首先，整合不同法领域的治理资源以强化算法规制。比如，可以通过引导平台优化算法打破“信息茧房”，以信息多样性促进理性对话，从源头净化网络生态、减少网络暴力发生。其次，在立法和执法过程中恪守比例原则，充分保障用户的合法权利和言论自由，避免过度限制言论。最后，发挥刑法作为保障法的底线支撑作用。面对日趋严峻的网络暴力态势，亟须完善相关犯罪的具体教义学解释，以提升刑法适用的精准性，为网络活动划定行为底线。

（三）刑法介入网络暴力的限度与新罪增设之非必要性

首先，刑法对网络暴力的介入应基于其功能定位进行审慎权衡，厘清网络的虚拟性与现实损害之间的关联性，以精准把握刑法介入的必要性及其限度。基于刑法的功能定位，刑法应在保障公民表达自由的前提下，发挥其惩治与预防网络暴力犯罪的功能。当网络暴力侵害受害者权益且造成超出民事救济范畴的损害后果时，刑法介入可提供更强有力的保护。然而，刑法介入须恪守罪刑法定原则，聚焦于破坏网络秩序的犯罪行为，警惕刑罚权不当扩张对作为网络空间基石的表达自由造成减损。

网络固有的虚拟性、匿名性与传播的迅捷性，亦增加了把握介入限度的难度，厘清网络行为与现实损害之间的内在关联，成为刑法有效介入的关键。如前所述，网络暴力并非囿于虚拟空间的符号冲突，而是可通过从虚拟至现实的伤害传递机制，对现实生活造成实质性侵害。网络行为与个体的现实身份、社会关系和社会评价紧密相连，网络暴力正是利用这种关联性将虚拟攻击延伸至现实生活，严重损害受害者的名誉、职业发展、人身安全乃至社会根基。因此，刑法介入应聚焦于网络行为所造成的现实损害后果，而非网络上的一般性言论冲突或失范行为。这要求立足于虚拟世界与现实世界相互作用、相互渗透的视角，准确研判网络暴力的本质和危害，从而将刑法规制重心聚焦于其现实损害。

其次，刑法对网络暴力行为之现实损害的介入，不应局限于个人法益视角，亦需兼顾其对社会秩序和公共利益的损害。网络暴力表现形式多样，损害程度各异，故刑法介入应避免“一刀切”式的简单规制，应针对不同的表现形式准确设定罪名。据此，可大致区分出人身保护模式和秩序保护模式两种类型：前者以个人法益为核心，聚焦网络暴力对个体的直接伤害，力求避免刑法的过度扩张；后者则将网络暴力视为对网络空间秩序的扰乱，更侧重于其对网络生态和公共利益的危害，旨在维护网络公共秩序[21](P105)。然而，需注意个人法益与超个人法益之间天然存在张力：一味侧重个人法益保护，可能导致打击力度不足，难以遏制网络暴力蔓延；过度强调公共利益，则可能导致打击面过宽。

因此，为避免刑法介入的泛化及其对言论自由的不当限制，宜引入对超个人法益的限缩机制，将生活世界和“系统”中不同类型的网络暴力加以区分，并结合行为对个人法益及社会秩序的损害程度，进行

类型化分析和精细化评价。在生活世界中,针对个人的网络诽谤、侮辱乃至线下暴力行为,应着重考量其对个人法益的损害程度,适当采取更为积极的干预措施。例如,网络教唆、煽动他人实施针对个体的线下暴力,其实际危害已超越网络空间,往往通过网上暴力与线下暴力的交织作用,对相关法益造成直接且严重的损害。此类行为已触犯旨在保护特定法益的现有举止规范,可直接适用故意伤害罪、寻衅滋事罪等罪名。对于散布虚假信息的网络诽谤行为,则需考量其是否对个人名誉权造成实质性损害。一旦损害程度轻微且并不涉及实质性的法益侵害,便应通过民事诉讼等途径解决,避免刑法过度介入。对于泄露隐私信息的网络暴力,同样需考量其对个人信息自决权的法益侵害程度,以准确认定是否构成侵犯公民个人信息罪。至于关涉"系统"层面的网络煽动、编造传播虚假信息和寻衅滋事等网络暴力行为形态,则须审慎评估其对公共秩序的实际威胁。生活世界中与个人交往利益相关的轻微网络摩擦,不宜轻易诉诸刑罚。应审慎适用寻衅滋事罪,并优先适用编造、故意传播虚假信息罪等特别罪名。

最后,考虑到现有刑法体系本身具备一定的规制能力,且司法实践中多数网络暴力行为可通过对现有罪名的解释适用得到有效处理,专门增设新罪的必要性值得商榷。关于应否专门增设新罪,有学者主张增设"网络暴力罪"等专门针对网络暴力行为的罪名以更精准地打击此类行为[22](P120),也有论者指出网络暴力行为情形复杂不适合直接入罪,可增设煽动仇恨情绪罪[23](P39)。

然而,在现有刑法体系相对完善的情况下,增设新罪名存在如下弊端:一是罪状难以精确界定。网络暴力行为样态繁杂且演变迅速,难以用固定的罪状予以概括。若罪状设置过于宽泛,易导致打击泛化、不当限缩言论自由;若罪状设置过于狭窄,则挂一漏万,难以全面规制。二是法益保护功能难以实现。如何周延保护法益是增设新罪的一大难题。网络暴力犯罪的首要规制目的在于保护个人法益,且传统刑法对个人法益的保护框架以造成实际损害后果为原型。而网络暴力的侵害链路复杂多变,增设新罪突破传统的实际损害后果模型,可能打破现有罪名体系的内在逻辑与协调性,带来入罪范围的无序扩张。三是增加司法适用难度。我国现有刑法体系已包含多个可适用于网络暴力案件的罪名,其解释与适用相对成熟,足以涵盖大部分网络暴力行为。若增设新罪,在侮辱罪、诽谤罪等罪名与新罪竞合时,则可能出现适用困局[24](P44)。针对侵害超个人法益的网络暴力行为,亦可援引帮助信息网络犯罪活动罪和非法利用信息网络罪等罪名进行规制。因此,既然通过扩张解释或立法完善现有罪名即可实现有效规制,则并无必要增设新罪。更务实的路径在于完善现有罪名的解释,如扩张解释拒不履行信息网络安全管理义务罪的适用范围,并夯实行刑衔接机制。

四、网络暴力刑法规制进路的优化

既然增设新罪并非网络暴力刑事规制的最佳路径,更为妥当的选择是,着重从解释论层面优化刑法规制进路。这要求充分结合网络暴力的生成机制,准确把握网络平台的核心枢纽地位与治理主体责任,有针对性地强化对平台和关键信息传播者的治理。同时,应以罪刑法定原则为根本遵循,通过对既有罪名的扩张解释和必要的立法完善,合比例地回应不同类型的网络暴力治理需求。

(一)网络平台的管理义务与刑事责任认定

如前所述,网络暴力的复杂生成机制既源于社交网络平台的结构性缺陷,也与现实冲突网络化的流量扩散链条密不可分。这意味着,局限于对个体施暴者的末端惩治往往收效有限,刑法规制重心应前移至对网络暴力生成之关键节点的结构性治理。据此,应强化平台责任和关键信息传播者责任,并审慎地扩张适用既有罪名。

首先,厘清社交网络平台的主体责任,是有效干预网络暴力生态链的首要环节。社交网络平台对信息分发、规则制定与用户行为均具有强大的塑造力,深度介入并形塑着网络交往秩序。它们既是信息流量的守门人,也是网络生态的构建者,对网络暴力的滋生、蔓延乃至恶性升级往往具有显著的结构性影响。凭借这种优势地位与控制能力,平台理应负有特殊的安全管理义务。强化这一平台责任,成为从源

头上遏制网络暴力的关键。

在这方面，《中华人民共和国网络安全法》《中华人民共和国个人信息保护法》等法律虽规定了网络服务提供者的安全管理义务，但相关规范较为零散，缺乏针对网络暴力这一特定治理对象的规制密度及体系化设计。《治理规定》等规范性文件则对此有所细化，将平台对网络暴力信息的预测、识别和及时处置等纳入其安全管理义务。追溯其法理基础，平台的信息网络安全管理义务植根于更深层次的法律原则乃至宪法要求，包含了其所应承担的法秩序维护义务与基本权利促成义务等公共义务[25](P19-21)。落实到网络暴力的具体治理，这一平台责任具体体现为：在技术治理与法律治理的交叉区间中发挥其独特优势，从前端预防、过程干预到后续修复，实现对网络暴力的全链条治理[26](P76)。当然，平台与用户之间在法律上系平等主体，其“私权力”属性的管理权限理应受到必要限制，聚焦于明显违法信息的甄别处置和高风险信息的前瞻性治理，而非对用户内容实施无差别、侵入性的全域监管。

具言之，可将网络平台的信息网络安全管理义务提炼为用户赋权、风险预警、信息干预和损害救济四个阶段，并据此进行义务细化。其中，用户赋权关涉增强用户的数字空间自主权这一有效治理网络暴力的逻辑起点。平台应建立用户自我防护机制，使之自主支配其专属空间[27](P39)，以实现风险的前端防控。相应的防护功能应尽可能充分，如屏蔽特定用户、设定信息可见范围、禁止内容转载或评论、调整私信接收权限以及自定义关键词过滤等用户选项，以减少潜在伤害；对于未成年人、社会弱势群体等特殊主体，还应另行配置差异化的保护机制。风险预警则要求治理防线前移。平台应借助大数据分析构建网络暴力风险预警体系，将事件敏感度、主体特征、参与规模、语义分析等多维度指标纳入统一的风险量化模型，以履行对网络暴力的早期识别与风险预防义务。一旦发现异常账号，应采取动态核验身份、弹窗警示及限制流量等干预措施，确保关键信息节点的责任可识别、信息流动的路径可追溯。信息干预的边界，则应以合理的内容审核与用户管理为限。为此，宜针对不同程度的网络暴力构建分级响应模型，将一般性言论争议、恶意网络暴力与重大网络暴力事件区别对待。对一般性言论争议，仅作用户提示或内容降权处理；对恶意网络暴力，则应采取封禁、限制传播等惩戒手段；而当出现大规模恶意攻击、严重侵犯个人隐私或煽动社会对立等重大网络暴力事件时，干预强度则应相应提升，并严防跨平台扩散。此外，平台还需完善损害救济，及时为网络暴力受害者提供救助服务。考虑到技术治理难免存在误判，还应建立完备的申诉与纠错机制，保障用户的言论自由等合法权益。

在此基础上，在刑事责任认定层面，对于未履行信息网络安全管理义务、导致网络暴力事件发生或危害后果扩大的平台，可探索适用拒不履行信息网络安全管理义务罪。应明确，平台基于民事、行政法规承担的安全管理义务并不直接等同于其刑法义务。平台的刑法义务界定应遵循必要、可能和可期待的比例原则，进行实质判断。其中，必要性体现在其措施必须具有治理网络暴力的实质效果，且对平台影响最小[28](P21)。例如，当被害人提供证据并就其所遭受侵害向平台发出明确通知后，平台经核实应及时采取必要的技术处置措施阻止侵权信息传播。若平台不作为导致违法信息大量传播或引发严重后果，且监管部门责令后仍不改正，则可能构成本罪。进一步地，仅当平台未正确履行其刑法义务时，才应承担刑事责任。根据对网络信息内容的分级管理，平台的刑法义务范围应与其对网络信息内容的甄别管理能力相适应，仅对法律明确要求处理且具备识别可能性的明显违法信息负有刑法上的作为义务。

面对网络暴力高发与平台治理重要性凸显的态势，亦有论者主张激活平台片面帮助行为的间接刑事责任，即平台若明知网络暴力信息存在而不作为，间接促成损害结果放大，则应通过片面帮助犯或共犯正犯化等模式追究其共犯责任[29](P147)。然而，该观点将间接因果关系与刑事责任简单关联，不仅理论根基薄弱，更导致平台义务过重，并模糊作为犯与不作为犯的界限。一方面，将平台的不作为等同于刑法意义上的帮助行为缺乏理论依据，不仅难以证明平台帮助他人实施特定犯罪的故意，亦与刑法中帮助行为通常要求对正犯行为具有物理或心理促进作用的标准相龃龉。另一方面，在刑法已为平台设定特定不作为刑事责任路径的背景下，将平台的风控失职等同于对具体犯罪的实质性促进，既有悖于共犯

理论的基本构造,也破坏了刑法体系的协调性。这会模糊平台自身管理责任与网络暴力直接实施者责任的界限,致使责任泛化。因此,对于具有口袋罪风险的帮助信息网络犯罪活动罪,应严格限制其适用范围。扩张适用拒不履行信息网络安全管理义务罪以追究平台的直接刑事责任,是更为妥当的追责路径,既符合罪责刑相适应原则,亦可避免刑法过度介入网络空间,更好地平衡网络言论自由与网络安全治理。当然,扩张适用该罪名时亦应严守其构成要件,特别是"经监管部门责令采取改正措施而拒不改正"的程序性前提以及"致使违法信息大量传播"或"造成严重后果"等实质结果要求,避免过度介入。

(二)网络暴力参与者的归责基础与刑事责任

以KOL与MCN机构为代表的关键信息传播者,作为信息生产和传播的枢纽,可通过议程设置、话语引导、粉丝动员等方式,在网络暴力事件中扮演催化乃至助推角色,故应将之纳入法律治理体系,并结合网络暴力的生成机制,厘清其责任的法理基础与具体边界。近年来,诸多网络热点事件的生成与发酵过程均可见关键信息传播者的身影,其如信息放大器般裹挟着用户,甚至通过制造对立、挑动情绪攫取流量,致使网络暴力周期拉长,波及更广,甚至从个人污名化演变为群体污名化[30](P30)。因此,鉴于其在信息传播链条中的特殊地位与影响力,KOL和MCN机构应当对相关法益的保护承担规范义务。其中,KOL作为内容创作者负有与其影响力相匹配的注意义务,须对所发布或转发信息的真实性、合法性进行审慎核查,避免传播煽动性、侮辱性、诽谤性或侵犯他人隐私等有害内容。而MCN机构作为内容产业的聚合运营主体,兼具准内容管理平台角色,其注意义务趋近于网络服务提供者,基于其对签约KOL的管理职能及其对内容分发渠道的控制力,应承担更高层级的内容管理与风险控制责任,建立健全对签约KOL的内容审核、监测与风险预警机制,阻断虚假信息和煽动性言论的规模化传播。

鉴于KOL与MCN机构在网络暴力事件中的角色及影响力差异,其责任设定亦应差异化。针对KOL,应综合考量其粉丝规模、过往行为及专业背景等因素,确保其注意义务与其言论的实际社会影响力及可预见的风险程度相匹配。MCN机构则在内容生产、流量管理及营销推广等方面均负有法律义务:事前应对签约KOL的内容进行严格审核与风险预警;事中应动态监控网络舆情并对网络暴力风险进行应急处置;事后则应积极配合司法机关留存证据与协助调查[31](P157)。此外,MCN机构还应履行流量管理义务,防止通过恶意营销制造或激化网络暴力,并杜绝利用水军等手段放大事件影响。

在此基础上,关键信息传播者的刑事责任认定需从主客观两个维度进行精细化考察。客观上,其行为须在网络暴力传播链条中发挥关键的媒介放大作用,如未经核实即转发或评论攻击性信息,或以流量扶持、话题炒作等方式扩大其影响力,致使事件升级为网络热点并造成严重后果,且该行为与严重后果之间存在刑法上的因果关系。主观上,其须具有故意,即明知相关信息虚假或具有严重人身攻击性,且预见到其传播极可能造成严重损害,仍积极追求或放任该结果的发生。对于MCN机构等主体,其履行内容管理职责时的不作为亦可能成立犯罪,此时需重点考察其是否违反法定的信息网络安全管理义务,以及不作为与损害后果之间的因果关系。

就刑事责任的具体认定路径而言,当前仅以侮辱罪、诽谤罪、侵犯公民个人信息罪等罪名规制关键信息传播者的严重违法行为,存在评价不充分的问题,尤其难以充分评价其利用网络影响力放大危害的行为以及MCN机构的不作为。因此,亟待在立法层面强化其不作为或管理疏失的责任形态。可借鉴现有网络安全管理规定中网络平台责任的具体内涵,拓宽拒不履行信息网络安全管理义务罪在网络暴力场景下的适用空间。可通过立法修正,明确以MCN机构为代表的关键信息传播者在网络暴力场景下的信息网络安全管理义务及违规认定标准,从而为该罪名的适用提供更为确切的规范指引。

与此同时,还须强化其他网络暴力参与者的刑事责任,尤其是参与制造虚假流量、恶意操纵舆论并从中牟利的组织或个体,以及为其提供技术支持、组织策划或具体实施操纵行为的网络黑灰产从业者。如前所述,雇佣水军进行恶意炒作、刷量控评等行为,其本质是通过技术手段人为扭曲信息传播格局,制造虚假民意,从而放大特定负面信息对被害人的攻击效应,加剧网络暴力蔓延并放大其法益侵害。因

此，刑法须严厉打击流量造假，将此类有组织地制造虚假流量、恶意操纵舆论的行为，视其具体手段与危害后果，适用非法经营罪、非法利用信息网络罪乃至寻衅滋事罪，并适当提高量刑标准。例如，为蹭炒热度、推广引流而利用公众账号传播网络暴力信息的行为，除适用诽谤罪、侮辱罪外，若符合“为实施诈骗等违法犯罪活动发布信息”等要件，亦可适用非法利用信息网络罪。

此外，鉴于人工智能驱动的机器人水军正成为流量操纵的重要力量，应将之视为网络犯罪的新形态，对其技术提供者、使用者均需追究相应的刑事责任，避免技术发展异化为网络暴力的帮凶。应加强对网络黑灰产业链的打击，重点惩治提供技术工具、数据服务及账号资源的上游环节，从源头上遏制网络水军的滋生。结合网络暴力的生成机制，个人信息的泄露是受害者网络身份和现实身份相关联、使网络暴力破坏力加剧的关键环节。因此，对相关网络黑灰产犯罪应充分适用侵犯公民个人信息罪，并以此作为阻断网络暴力风险的源头性举措。对恶意传播网络暴力信息的个体行为，则应综合传播情境、范围、手段、后果和主观恶意等因素，酌情适用侮辱罪、诽谤罪或寻衅滋事罪，避免打击范围不当扩张。

（三）累积犯之否定与既有罪名的扩张适用

针对网络暴力的群体性和复杂性带来的法不责众难题，有学者提出累积犯的规制主张，但该主张忽视了网络暴力的复杂生成机制，可能导致规制失焦与刑罚泛化。不同类型的网络暴力的参与行为存在显著的异质性，将其一概而论缺乏依据。除通过共同犯罪对关键参与者予以规制外，还可通过侮辱罪、诽谤罪、寻衅滋事罪等既有罪名的扩张解释或立法修正，实现对严重网络暴力行为的充分评价。

首先，将网络暴力等同于累积犯的观点，在刑法教义学和刑事政策层面均有待商榷。有观点认为，网络暴力源于不同网民海量行为的“真实累积效应”，产生网络失范行为的累积性侵害，此类累积危险行为属于累积犯[32](P30)。然而，累积犯的典型适用场景是对重大集体法益的提前保护，而网络暴力则属于主要集中于个人法益的犯罪类型的集合，二者的适用场景存在根本差异[28](P15-16)。将网络暴力简单归类为累积犯是对其本质的误解，不仅可能导致规范目的错位，更有损及公民言论自由之虞。诚然，网络暴力在外观上具有参与者众、行为分散、因果链条复杂交织等与累积犯相似的表征，但若仅凭此表象便将所有参与者不加区分地视为正犯，则忽视了个体间的罪质差异，有悖于归责原理。从刑法教义学角度看，既然网络暴力的危害后果常由群体互动、平台机制、社会情绪等多重复杂因素的共同作用所引发，离散个体与最终结果之间的因果关系难以精确证立，将数量庞大、行为贡献差异悬殊的个体悉数纳入归责范围，缺乏客观归责层面的根据。在主观归责层面，更难以认定所有参与者具备共同的犯罪故意。网络空间系意见表达与信息流通的重要场域，部分言论即使带有情绪化色彩或批评性意见，只要未逾越法律边界，便不具有可谴责性。将所有相关言论的发表者均推定为具有犯罪意图，既有违主客观相统一原则，也与罪刑法定原则所要求的明确性与可预测性相悖。从刑事政策角度看，将主观动机、行为样态以及对最终危害结果的贡献程度迥异的参与者不加区分地归责，不仅严重违反罪责刑相适应原则，更易引发钳制言论自由的负面效应，压制公共讨论活力，最终损害网络生态。

其次，刑法规制不应简单考虑参与人数多寡，而应精准甄别不同参与者的角色和责任，着重规制其中起关键作用的主导者和煽动者。结合网络暴力的生成机制，网络暴力事件的迅速发酵与实害化，核心症结即在于少数主导者或煽动者对平台算法与注意力机制的恶意利用。此类主体往往蓄意捏造、散布关联被害人真实身份的虚假信息或极端言论，再借由操纵信息流、煽动群体情绪、设置议题陷阱等手段收割流量并激化矛盾，最终形成对受害者的集中攻击。据此，应聚焦于识别并惩处此类网络暴力事件中的组织、策划及核心推动者。事实上，《指导意见》已初步体现出区分恶意发起者、组织者、恶意推波助澜者及屡教不改者的规制思路，但各类主体的行为性质与责任形态应如何精确界定，仍有待深入阐释。

相较于累积犯评价路径的整体性与模糊性，区分正犯与共犯的理解显然更契合网络暴力的刑法规制。基于这一认定思路，恶意发起者作为网络暴力的始作俑者，多出于诽谤、侮辱、侵犯隐私等不法目的，率先发布虚假、侮辱性或涉隐私信息挑起事端，其行为兼具主动性、指向性与煽动性。因此，可根据

具体情节将之认定为侮辱罪、诽谤罪、侵犯公民个人信息罪等的直接正犯。组织者在幕后策划、指挥、协调或资助,为网络暴力提供组织保障并对事件发展起关键推动作用。此类行为已超出帮助范畴,对犯罪进程具有实质支配力,可能构成侮辱罪、诽谤罪、寻衅滋事罪或非法利用信息网络罪的共同正犯或间接正犯。恶意推波助澜者的情形相对复杂。此类行为人明知信息严重失实、涉嫌侮辱诽谤或已造成严重伤害,仍通过转发、评论、制造新话题、提供攻击素材等方式积极参与,显著加重了法益侵害。由于其行为已超出一般性参与范畴、体现出对不法行为的认同与促进意图,应根据其参与行为的具体方式、强度、主观恶性程度以及与核心行为的关联性,审慎判断其是否构成侮辱罪、诽谤罪的共同正犯或帮助犯。特别是MCN机构等关键传播节点,若其为追求流量或商业利益而故意实施此类行为,更可探讨非法经营罪、非法利用信息网络罪等秩序性犯罪的适用可能,从严审视其刑事责任。针对实施网络暴力行为的屡教不改者,若其行为已构成犯罪,鉴于其主观恶性较大,在刑事责任认定时应依法从重处罚。

至于数量庞大的一般参与者,其行为多属偶发性、情绪化的跟帖、点赞,主观上多无明确的加害意图,客观上对法益侵害的贡献也相对有限,通常难以达到相关犯罪“情节严重”的入罪标准。对其刑事责任的认定,需综合考量行为类别、对象、参与人数、信息内容、发布频次与传播场景等因素,进行个案判断[33](P96)。若不构成犯罪,可视情节施以行政处罚或批评教育;情节轻微且未造成实质损害的,则无需追究法律责任。

再次,网络暴力的刑法规制应恪守罪刑法定原则,通过激活既有罪名实现对严重网络暴力行为的精准打击。网络暴力本质上系一种高度同质性群体对少数个体的集体恶意攻击,甚至可能导致受害者自杀等严重后果。据此,解释论上不应排除基于对实行行为和教唆自杀行为的个案实质解释,从而适用过失致人死亡罪、故意伤害罪乃至故意杀人罪的可能性。至于网络暴力中常见的道德审判行为,一旦情节严重,同样可适用侮辱罪或寻衅滋事罪。针对“开盒挂人”等行为,即便部分信息由受害者自行公开,也不能当然阻却行为的违法性。若行为人未经授权系统性地搜集、整合、关联这些信息,特别是深度挖掘并将足以识别特定自然人身份或活动轨迹的敏感个人信息公之于众,用于煽动、组织网络围攻或线下骚扰,实质侵犯公民的个人信息自决权,且符合“非法获取”或“非法提供”等侵犯公民个人信息罪的行为特征,达到“情节严重”标准即可入罪。同时,平台负有审核、移除此类侵权信息的法定义务,理应建立有效的技术识别与用户举报处理机制。倘若其系统性地怠于履行信息内容管理责任致使“开盒”信息泛滥,可能构成拒不履行信息网络安全管理义务罪。

寻衅滋事罪在网络暴力案件中的扩张解释边界亦亟待厘清。该罪的构成要件之一是“在公共场所起哄闹事,造成公共场所秩序严重混乱”。一些网络暴力事件不仅严重侵害公民的行动与意思活动自由,更严重扰乱网络空间乃至线下社会秩序,造成恶劣社会影响。对于此类行为,适用寻衅滋事罪兼具规范依据与现实合理性。相关司法解释亦试图通过界定网络上“起哄闹事”与“造成公共秩序严重混乱”等标准,扩张适用本罪。然而,即便基于客观解释论肯定本罪的扩张解释空间,仍应将入罪标准严格限定为对公共场所秩序的实质侵害,并结合网络暴力的具体作用场景,谨慎平衡言论自由与秩序法益。

网络场景并不当然构成网络公共场所,只有具备公共讨论功能、面向不特定或多数人开放的网络平台或社群,才符合公共场所的文义范围。例如,微信群聊虽具网络公共场所的潜在属性,但基于亲属或特定身份组建的小微群组,则因其具有私人性而应被排除在外。而所谓侵害“网络公共场所秩序”,则应与一般性网络争执或情绪宣泄相区别,特指有组织、大规模、持续性的网络围攻。其典型情形包括操纵“水军”控评淹没正常讨论、恶意刷屏瘫痪公共交流空间、散布虚假信息引发群体性恐慌或对立等。此类行为不仅侵害特定个体,更通过制造网络空间的混乱、恐慌与对抗,严重破坏特定网络交往场景下特定网络社区的正常运行,进而动摇不特定多数用户对网络空间交往的安全期待,且该风险可经由现实身份关联将伤害延伸至受害者的现实生活,足以引发线下秩序的连锁反应。因此,必须综合考量网络公共场所的性质、起哄闹事的参与情况与具体情境、后续线上线下影响的范围与程度等因素。即便存在言论冲突,唯有当行为明显逾越言论自由的正当边界,存在“起哄闹事”的严重行径,造成特定网络公共场所秩序严重混

乱并严重危及受害者现实安宁，且行为人主观上具有恶意挑衅、无事生非等动机时，方可审慎适用本罪。

最后，针对将现有罪名适用于网络暴力案件时面临的现实挑战，亟须在立法层面予以完善。其一，诽谤罪的适用存在争议。诽谤罪要求行为人“捏造事实”诽谤他人，而网络暴力言论虽可能包含负面评价，但其内容往往并非完全虚构，而是基于特定事件的道德审判与跟风指责，此时是否契合“捏造事实”的构成要件尚存分歧。其二，侵犯公民个人信息罪的适用亦面临挑战。适用争议主要体现于“人肉搜索”行为，即通过整合已公开的公民个人信息揭示他人真实身份的定性问题。除解释学路径外，立法修正有助于对这些认定挑战一锤定音：可审慎研究是否需要调整诽谤罪的构成要件表述，将“捏造事实”完善为“捏造、恶意歪曲事实或进行严重误导性陈述”等情形，以精准涵盖网络暴力中常见的基于选择性信息或歪曲解读而发起的名誉攻击行为；还应将非法利用个人信息的行为纳入侵犯公民个人信息罪的范围，对于恶意整合、关联分析并公开散布可识别特定个人信息的行为，即使部分原始信息曾被公开，但若该整合、散布行为超出了合理使用范畴并对个人信息权益造成严重侵害，亦应在立法中明确其可罚性，以明确网络暴力情境下公开散布他人个人信息的刑事责任，增强本罪与网络暴力场景的适配性。

网络暴力作为一种依托社交网络平台、借助流量传播的新型违法犯罪形态，其治理并非传统法规范的简单移植，应在厘清其生成机制的基础上，审慎探求领域立法和刑法介入的恰当路径。准确把握网络暴力的生成机制，需明确的结构性现实是，社交网络平台的技术架构与商业模式并非中立，其算法驱动、流量至上等特征在为用户提供便捷的同时，也为网络暴力滋生蔓延提供了结构性土壤。同时，网络空间并非世外桃源，社会焦虑、群体极化等现实问题的网络映射，以及网络暴力与现实身份的关联性，共同构成了网络暴力治理的复杂图景。反思既有的网络暴力治理进路，技术治理模式虽不可或缺，但也可能带来对公民权利和言论自由的侵犯，应推动其与法律治理相互补充，并进一步优化法律治理模式。应构建以领域立法为基石、跨领域协同为保障的综合治理体系。就刑法规制而言，更有效的路径在于强化社交网络平台与关键信息传播者的责任约束，优化拒不履行信息网络安全管理义务罪、侮辱罪、诽谤罪、寻衅滋事罪等罪名的解释适用，并有针对性地对现有罪名进行立法修正，避免新罪增设可能带来的过度规制。未来仍待研讨，如何在法秩序统一的视野下，借由法域衔接的完善，更清晰地划定刑法介入的边界。

参考文献

[1] 柳思思.网络语言暴力问题研究.北京：人民日报出版社，2018.

[2] Jamie Susskind. *Future Politics: Living Together in a World Transformed by Tech*. Oxford: Oxford University Press, 2018.

[3] 胡泳.全球开放互联网的歧途.太原：山西人民出版社，2023.

[4] Cass R. Sunstein. *#Republic: Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press, 2018.

[5] 赫克托·麦克唐纳.后真相时代.刘清山译.北京：民主与建设出版社，2019.

[6] 艾伯特-拉斯洛·巴拉巴西.链接：商业、科学与生活的新思维（十周年纪念版）.沈华伟译.杭州：浙江人民出版社，2013.

[7] Hamid Jahankhani. *Cyber Criminology*. Cham: Springer, 2018.

[8] 雪莉·特克尔.群体性孤独：为什么我们对科技期待更多，对彼此却不能更亲密？周逵、刘菁荆译.杭州：浙江人民出版社，2014.

[9] 南希·K.拜厄姆.交往在云端：数字时代的人际关系.董晨宇、唐悦哲译.北京：中国人民大学出版社，2020.

[10] 欧文·戈夫曼.日常生活中的自我呈现.冯钢译.北京：北京大学出版社，2008.

[11] 赵熠如.多平台宣布实行“大V实名制”.中国商报，2023-11-02.

[12] 任秀，王咏梅.全网实名制后的困境与对策研究.湖北社会科学，2020，(4).

[13] 单民，陈磊.博弈与选择：以实名制遏制网络言论犯罪的可行性分析.河北法学，2015，(9).

[14] 张凌寒.“不良信息”型网络暴力何以治理——基于场域理论的分析.探索与争鸣，2023，(7).

[15] 王燃.论网络暴力的平台技术治理.法律科学（西北政法大学学报），2024，(2).

[16] 李晟.国家安全视角下社交机器人的法律规制.中外法学，2022，(2).

[17] 周佑勇.从部门立法到领域立法:数字时代国家立法新趋势.现代法学,2024,(5).
[18] 刘艳红.网络暴力治理的法治化转型及立法体系建构.法学研究,2023,(5).
[19] 彭诚信.领域法学视野下的数字法问题.政法论丛,2024,(1).
[20] 罗翔.网络暴力刑法规制的路径选择及反思——以侮辱罪的拆分为切入.中外法学,2024,(2).
[21] 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.法律科学(西北政法大学学报),2023,(5).
[22] 陈罗兰.网络暴力的刑法治理与罪名增设.法律科学(西北政法大学学报),2023,(5).
[23] 田宏杰.网络暴力刑法治理中的"法不责众"困境及其化解.法学杂志,2024,(1).
[24] 储陈城.刑法应对网络暴力的流变及其基本立场.中国刑事法杂志,2023,(4).
[25] 陈斌.数字平台义务创设的宪法基础.环球法律评论,2023,(3).
[26] 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.南京大学学报(哲学·人文科学·社会科学),2024,(1).
[27] 刘文杰.网络暴力中个体维权困境与平台安全保障义务.国家检察官学院学报,2023,(5).
[28] 李本灿.网络暴力犯罪处罚范围的教义学建构.环球法律评论,2025,(1).
[29] 龚文博.论网络暴力治理中的平台间接刑事责任.法制与社会发展,2024,(6).
[30] 王静.数字公民伦理:网络暴力治理的新路径.华东政法大学学报,2022,(4).
[31] 刘艳红.MCN机构作为网络暴力新型责任主体的地位确立与义务设置.东方法学,2025,(1).
[32] 冷必元.群体性网络暴力累积危险行为的刑法治理.法治研究,2023,(5).
[33] 于冲.网络"聚量性"侮辱诽谤行为的刑法评价.中国法律评论,2023,(3).

On the Generative Mechanisms of Cyber Violence And The Improvement of Criminal Law Regulation Approaches

Ma Yongqiang (Jilin University)

Abstract　An in-depth exploration of the generative mechanisms of cyber violence serves as the logical starting point and practical foundation for its effective regulation. Fundamentally, the generation of cyber violence is driven by multiple factors, including structural deficiencies in social networking platforms, the online projection of real-world social conflicts, and the linkage between cyber violence and offline identities. A critical review of existing governance approaches reveals that technological governance model, indispensable as it is, is by no means a fundamental solution, as its efficacy and potential risks warrant careful assessment; strengthening the legal governance model is therefore the key to ultimate solution. To this end, a comprehensive governance system should be constructed based on the generative mechanisms of cyber violence, with sectoral legislation as the cornerstone and cross-sectoral coordination as the safeguard, while clarifying the limits of criminal law intervention and exercising due caution in establishing new offenses. To optimize the approach of criminal law regulation, efforts should be grounded in the particularities of cyberspace governance: the criminal liability of online platforms should be reinforced through the application of criminal law; the accountability of key information disseminators and other participants should be thoroughly examined; it is necessary to reevaluate the application of cumulative offenses and anchor the approach firmly in the principle of nullum crimen sine lege, and the regulatory capacity of criminal law will be substantially enhanced through expansive interpretation of existing offenses and necessary legislative improvements.

Key words　cyber violence; platform responsibility; sectoral-specific law; cybersecurity management obligations; doxxing; criminal law interpretation

■ 作者简介　马永强,吉林大学理论法学研究中心研究员,吉林大学法学院副教授,吉林 长春 130012。
■ 责任编辑　李　媛