

走向媒介正义：网络暴力信息治理的平台角色与法治实现

朱笑延

(吉林大学法学院, 吉林长春, 130012)

摘要：网络暴力信息治理义务的增设，标志着平台成为治理网络暴力的“超级责任体”。洞悉网络暴力信息在平台环境中生成、扩散、异化的内在规律，成为理解数字时代网络暴力法律规制转型的关键。引入传播学的数字媒介理论，可以发现网络暴力信息并非由“乌合之众”自发聚集而成，而是在平台鼓励偏激言论、放大极端情绪、助推舆论集群的媒介不公环境下滋生的。搭建优化内容质量、鼓励理性讨论、提供被害救济的媒介正义秩序，成为平台履行网络暴力信息治理义务的独特内涵。平台应在事前预警环节设置与自身内容生态相适配的内容监测机制，实时感知网络暴力信息风险；在事中处置环节搭建健康引导网络舆情走向的议程框架，有效阻断网络暴力信息扩散；在被害保护环节配置具备及时响应能力的技术反制措施，防止被害人遭受网络暴力信息的舆论宰制。

关键词：网络暴力信息；平台义务；媒介正义；内容合规

中图分类号：D922.16

文献标识码：A

文章编号：1672-3104(2024)01-0050-13

一、问题的提出

面向平台设置专门化的网络暴力信息治理义务，成为数字时代网络暴力法律规制的全新走向。2022年11月，网信办印发的《关于切实加强网络暴力治理的通知》首次提出压实平台在网络暴力信息治理中的主体责任，并初步勾勒了平台在网络暴力信息识别、预警、处置等环节的治理角色。2023年7月，网信办发布的《网络暴力信息治理规定(征求意见稿)》(以下简称《征求意见稿》)更是围绕账号管理、信息发布、监测预警、举报救助、信息处置等治理环节，为平台量身定制了大量专门化、创新性的网络暴力信息治理要求。随着规制靶向、治理重心的迁移，平台实质

上被法律打造为网络暴力信息治理的“超级责任体”^[1]。然而，《征求意见稿》仅仅是从事前、事中、被害预防等几个方面为平台规定了框架性的义务履行要求，尚未明确具体化的义务履行标准。并且，平台对网络暴力信息的管控还牵扯网络内容创作自由、网络空间生态治理、被害人权益保护、平台主体治理责任等诸多复杂因素，导致平台的义务履行难度大大增加。因此，平台应该如何履行网络暴力信息治理义务，成为数字法学研究和网络暴力治理实践亟待解决的重要问题。

尽管平台的治理角色愈发重要，但网络暴力法律规制的相关研究长期缺乏平台的分析视角。既有研究主要从网络侮辱诽谤行为的刑法规制入手，分析网络暴力的刑法定性^[2]、网络暴力公诉机制的适用范围^[3]、诽谤罪“情节严重”的判

收稿日期：2023-08-11；修回日期：2023-09-22

基金项目：国家社科基金一般项目“数据正义视阈下犯罪的技术治理均衡发展研究”(20BFX066)；中国博士后科学基金第72批面上资助项目“中国社会的数字化转型对诈骗罪基本原理的再塑造研究——基于大数据样本的人工智能分析”(2022M721307)

作者简介：朱笑延，男，辽宁昌图人，吉林大学法学院“鼎新学者”博士后，吉林大学司法数据应用研究中心研究人员，主要研究方向：数字法学、中国刑法学、法学实证，联系邮箱：zxy666@jlu.edu.cn

断标准^[4]、网络暴力致人自杀死亡的刑事责任^[5]、网络暴力的刑法罪名增设^[6]、网络暴力法律规制的行刑衔接等内容^[7]。基于刑法教义学的精耕细作虽然贡献了网络暴力研究主要的知识增量，但也在无形中固化了“网络侮辱诽谤行为—刑法事后规制”的研究线条。随着《征求意见稿》搭建的面向平台治理的全新规制架构和独特义务类型，网络暴力的规制重心由单一的网络侮辱诽谤行为拓展至平台的网络信息内容生态，网络暴力的规制时点由事后规制转向了事前、事中干预，网络暴力的规制手段由被动的刑罚制裁升级为主动的预防治理。在根本性的规制变革和治理变迁之下，长期聚焦于网络侮辱诽谤行为的刑法规制分析已经出现文不对题、解释乏力的式微状态，平台已然成为网络暴力法律规制研究不可回避的理论要素。

值得肯定的是，近期的一些研究已经关注到新型网络犯罪治理中平台义务的大量增长，但聚焦于平台守门人规则的宏观理论研究忽视了网络暴力信息管控这一具体的内容治理场景。这些研究尝试从公私合作模式^[8]、数字看门人^[9]、公共责任公司^[10]等视角诠释平台义务中“国家管平台—平台管用户”的规制架构，探索建立有效的事前内容合规^[11]、完善主动的事中监管合规^[12]、强化平台治理绩效的政府监管等网络暴力平台治理的义务履行标准^[13]。但是，这些研究只看到平台义务一般性的规制架构，忽视了网络暴力信息这一特殊的平台内容治理场景。相应的义务履行标准仅具有方向性的参照意义，难以在网络暴力信息治理的独特场景中释放针对性的指导功能。

针对上述不足，基于“网络暴力信息+平台内容治理”，需要在以下几个方面推进既有研究：第一，为网络暴力的法律规制研究注入平台视角。不同于既有研究将网络暴力治理简化为规制网络侮辱诽谤行为的固有范式，本文将网络暴力问题视为数字时代独特的“平台现象”，希望从网络暴力信息在平台生成、扩散、异化的新视角重新解释网络暴力现象，及时回应由平台义务撬动的网络暴力规制思路变迁与治理逻辑转型。第

二，提炼网络暴力信息治理义务独特的内容治理逻辑。不同于既有研究将网络暴力信息治理义务视为一般性的看门人规则，本文聚焦网络暴力信息这一平台内容治理的场景特殊性，在充分理解平台环境中网络暴力信息生成、扩散、异化的数字化规律基础上，精确提炼网络暴力信息治理义务的规范性原理，以期指引平台对网络暴力信息治理义务的具体履行。

需要承认的是，由于平台中网络信息内容的生成、扩散、异化是传统法学研究没有深度涉足的领域，仅仅依靠法学的内部逻辑分析很难处理这项极富挑战性的工作。有鉴于此，本文希望借鉴传播学中的数字媒介理论，关注平台作为汇聚、整合、影响网络暴力信息的媒介性特质，在深度理解平台如何影响网络暴力信息传播的前提下展开分析：首先，运用数字媒介理论，理解平台在网络暴力信息生成、扩散、异化过程中的媒介作用力；其次，根据平台的媒介作用力，勾勒平台在网络暴力信息治理中的媒介化角色，以此重新阐释网络暴力信息治理义务独特的理论内涵；最后，根据网络暴力信息治理义务的理论内涵，探索平台有效履行此类义务的技术路径和实现方式。

二、制造媒介不公：网络暴力信息传播的平台视角

数字时代高度复杂的网络暴力问题可以从平台的媒介影响力视角获得新的解释。在数字媒介理论看来，平台作为用户生产网络信息内容、接收新闻消息、参与公共议题讨论的媒介型数字基础设施^[14]，既成为人类现实生活在数字空间的延展^[15]，又在无形中塑造着内容创造、新闻评论、公共参与的方式^[16]，进而对网络暴力信息生成、扩散、异化的全生命周期产生了结构性影响。具体而言，平台虽然以“打造包容、开放、互助社区氛围”的话语体系和价值隐喻构建自身形象^[17]，但在呈现、加工和聚合网络信息内容的媒介实践中，搭建了鼓励偏激言论生产、放大极端情绪扩散、助推舆论走向异化的媒介环境。因

此,网络暴力信息并非由“乌合之众”自发聚集而成,而是在平台“媒介不公”环境的诱导下逐步滋生的。

(一) 鼓励偏激性的内容生产

网络暴力信息的大量生产并非单纯是用户的个体化行为,而是存在平台这一数字媒介的隐性作用力。在“内容为王”的互联网竞争环境下,平台在不断刺激用户进行数字劳动和内容创作的同时,搭建了一个鼓励偏激性言论的公共讨论空间。随着攻击性的言论中伤在平台司空见惯,友善的交流辩论反而成为另类^[18],平台搭建的公共讨论空间逐步偏离了开放、宽容、互助的互联网原初精神,引发了鼓励敌视、压缩宽容的媒介不公问题,网络暴力信息正是在这样的特殊媒介环境下不断繁衍的。

首先,平台内容创作的低门槛引发了用户恣意发言、偏激表达而无所顾忌的媒介不公问题,提供了有利于网络暴力信息滋生的媒介环境。在数字媒介理论看来,网络信息内容的生产秩序并非由作为社会行动者的网民自发协作而生,而是由追逐商业空间的平台精心谋划而成^[19]。平台内容创作的低门槛虽然能够实现刺激用户内容创作、堆高流量收益的市场化目标,但也向用户传递了一种低成本发布偏激言论的可供性感知。为了支持大量社交媒体粉丝能够随时入场探讨^[20],平台设置了相对便捷、简单的用户注册审核机制,用户几乎只要使用智能手机“一键注册”,即可在社交媒体畅所欲言。但是,简易注册也使账号与用户本人在现实社会中的人格形象和社会声誉脱钩^[21],削弱了用户进行内容创作的社会责任感和在场感^[22],使得用户的发言几乎无所顾忌^[23],部分用户甚至会注册大量小号参与对喷、骂战,从而催生出职业性发布偏激言论、提供有偿“带骂”服务、批量化生产网络暴力信息的“网络水军”。而与前端的简易注册相呼应的则是后端的轻缓处罚,对于发布偏激言论的用户行为,平台虽然有一定的社区公约、账号处罚等手段予以限制,但这些处置多体现为扣除信誉积分、暂时禁言等轻缓化的处罚措施,其首要目的在于留住用户,而不是有效管控偏激性言论。这就导致

屡屡被惩处、扣除信誉积分的账号仍然能够不受限制地发布网络暴力言论。总之,平台过度刺激用户内容生产的运作机制,也在无形中诱发了偏激性言论屡禁不止的媒介不公问题,成为网络暴力信息大量涌入的重要因素。

其次,平台靶向化的话题讨论特质虽然在用户兴趣和关联话题之间建立起长期的情感联系,却也放大了用户敌视陌生话题的媒介不公问题,同样能够引发网络暴力信息的大量滋生。基于提升用户黏性、解放用户生产潜力、放大流量收益等动机,平台会记录符合用户偏好或关注的话题题材^[24],并通过算法机制向用户集中、精准推送类似的内容产品,从而将平台的内容特色与用户的个人偏好建立起巨大的关联黏性,以期在争夺用户资源、抢占流量高地的逐项竞争中取得优势。然而,平台对于同质性话题的打造,也在逐步塑造用户的选择性理解、选择性记忆的心理机制^[25]。在“只看到自己想看到的”信息茧房中,用户虽然强化了对自身偏好或关注话题的理解,但也降低了对陌生话题、不同群体的包容与认同,在遇到与自身价值立场、情感寄托不符的网络信息内容时,就极易产生语言暴力倾向。最为典型的案例就是不同平台对男女对立问题的不同立场。虎扑体育主要以男性用户为主,该平台对“出轨女”“高彩礼”“女拳”等话题的持续打造虽然能够博得男性用户的情感认同,但也极易激发男性用户对女性群体的偏见和歧视,从而在两性话题、恋爱话题、婚姻话题中滋生基于“厌女症”的性别语言暴力。与之类似的则是女性用户高占比的小红书,该平台对“凤凰男”“渣男”等标签话题的持续运营虽然能够吸引大量的女性用户入驻,但也容易降低女性用户对涉男性新闻的容忍阈值,“恐男”情绪的蔓延已然证明,平台靶向化的话题讨论机制更有利于提供鼓励敌视、制造偏激的交往场景。

(二) 加速极端化的内容扩散

网络暴力信息的大面积扩散并非是网络空间的偶然现象,而是受到平台流量化公共讨论机制的隐性推动^[26]。在流量至上的商业变现模式下,平台按照流量进行内容分发的运营模式虽然

显著增加了争议事件的曝光度和讨论量，但也将网络信息内容中最为尖锐、极端的网络暴力信息“置顶”传播。由此，新的媒介不公问题便在平台的流量化逻辑下出现——极端化的情感反应明显压制理性化的交流探讨，片面性的内容呈现逐步遮掩事件冲突的真实面貌。

首先，平台的流量逻辑会形成充斥极化情绪和语言暴力的网络舆情，感性的情绪宣泄推动了网络暴力信息的迅速扩散与二次生产，而理性的事件讨论则被压制在网络空间的阴暗角落，陷入“沉默的螺旋”。大多数平台会依据点赞及回复的数量高低对争议事件的舆情反应进行系统排名，市场化的媒介动力机制虽然显著提升了网络信息内容的传播效率，却也导致非理性的网络暴力信息因眼球效应而经常被置顶，理性的思辨和多元的表达下沉至海量评论的末端^[27]。理性讨论优于语言暴力的社会共识，在流量化的影响力赋权过程中被倒置。由于极端化情绪始终盘踞在头部版面，被置顶展示，受众的表层情绪体验逐级攀升，理性内容讨论维度和对话质量明显下降。于是，争议事件中的网络暴力信息迅速抢占了受众有限的注意力，在流量的加持下占据平台的头版头条，放大与流量规模相匹配的负面影响。例如2022年发生的“上海女子打赏骑手200元被网暴案”中，平台流量化的内容呈现逻辑将“善良这么廉价吗”“双十一快递买了100多个，结果给人家200”等获得上百条点赞的情绪化评论置顶宣传，并持续获得点赞、评论、转发的聚量效果，而为女子声援的评论内容和理性发言则被淹没在聚众讨论中。此外，流量化的逻辑还在信息扩散过程中造成网络暴力信息的二次生产。为了在对喷、互骂、引战过程中获得更多的回复讨论量，用户在参与公共事件的讨论中具有强烈的语言暴力意愿，希望通过极端的情感宣泄而不是理性的讨论参与来获得流量的青睐。部分“大V”更是将关注度放在了话题吸引力上，通过极端化的话语表达和情感宣泄吸附话题受众、抢占评论资源，导致争议事件的讨论进一步走向失控。例如2023年发生的“武汉小学生母亲坠亡案”，在丧子母亲被网民质疑过度索要赔偿之后，大量营

销号又采用“妆容精致”“三天换两双香奈儿鞋子”的标题，制造被网暴母亲丧子之痛与有精力打扮之间的道德冲突，刻意诱导网络暴力信息在该话题中二次集结，从而导致这一事件的语言暴力规模和社会影响迅速升级。

其次，平台的流量化逻辑还在于对争议事件进行片段化加工，将多元的冲突、复杂的流程置换为刻板的标签与简单的对立，诱使极端化的语言暴力和情绪反应进一步扩散。在注意力经济的驱动下，平台通常以简短的话题标签、热搜标题、关键词等形式，将多元的冲突、复杂的流程置换为刻板的标签与简单的对立。在平台“制造偏见”的内容加工逻辑下^[28]，用户的理性思考因信息内容的有效供给不足而退场，感性判断因数字媒介的实时在场渲染而爆发^[29]，带有大量情绪化色彩的网络暴力信息也就经过群体性的点赞、评论、转发，在高速扩散的过程中支配了整个争议事件的话题基调与情感架构。并且，平台内部本就积累了一些与网络暴力问题相关的对立话题，使争议事件的复杂冲突极易在平台的讨论环境下与官民对立、劳资关系、师生关系、性别歧视等敏感问题产生联系，导致大量关注这些话题的用户迅速涌入相关话题的探讨，并在极端化的情绪反应中完成网络暴力信息的二次生产。这就不难理解，为何网络暴力信息涉及的冲突事件虽然是新鲜的，但内容却始终在有限的话题和类似的情绪中反复循环^[30]。如2023年发生的“女研究生网暴大叔偷拍案”中，研究生张某质疑大叔偷拍的内容表述因为同时触及了两性、偷拍、女性权利保护等极易引发观点对立的平台争议话题，迅速吸引了平台受众对偷拍的道德排斥和对张某的情感共鸣，这些情绪宣泄无视了大叔没有偷拍的事实，径直将大叔刻画为“猥琐”“偷拍”“渣男”的刻板形象。总之，在流量化的内容呈现和加工逻辑下，平台实质上促成了网络暴力信息的快速扩散以及二次生产，明显压制了理性讨论应有的生存空间。

（三）助推单方面的舆论压制

网络暴力信息中针对被害人的舆论讨伐并非完全由“乌合之众”自发组织，而是需要嫁接

在平台的媒介性连结中方能形成压倒性的群体同盟与舆论权威。平台提供的媒介性连结所形成的“群体性力量”未必带来公共价值的实现,反而可能在加害人与被害人之间制造侵权易、救济难的媒介不公问题。诸如私信轰炸、人肉搜索、线下骚扰等数字暴力奇观,皆是媒介不公问题的具体表现形式。

从加害人的角度来看,平台提供的媒介性连结在表面上中立,实质上有利于网络暴力的加害人获得相对被害人的数字临时性权力,加害人由此获得了肆意践踏被害人人格尊严、完全剥夺被害人数字生命的舆论权威。网络暴力信息数量上的增加、传播力的扩展,实质上代表着语言暴力的道德指向获得了平台之中网民群体的认同,意味着加害人在平台媒介环境中单方面掌握了临时性的舆论话语权,获得了讨伐被害人的道德基础。在由平台媒介搭建、由施暴群体主审的数字审判席上,加害人能够对被害人形成强大的群体舆论攻势,抹杀被害人的数字公民身份,否定被害人的数字化生存权利,运用各种表达方式意图提前宣告被害人的“社会性死亡”^[31]。如果基于语言暴力的“对人不对事”取代了基于事件讨论的“对事不对人”^[32],那么网络暴力信息在平台中的扩散就绝非侮辱诽谤言论数量的单纯叠加,而彻底异化为道德审判的情感政治^[33]。

面对加害人单方面的舆论压制,平台复杂的网络信息传播机制却不利于被害人展开权利救济。深陷于网络暴力信息泥潭的受害人既无法获得公众理解^[34],又没有能力矫正损害。在平台搭建的公共讨论空间中,网络暴力的加害人找到被害人极其容易,但被害人寻找加害人却极为困难。即使刑法第246条侮辱罪、诽谤罪赋予了被害人自诉的权利救济渠道,但在平台复杂的传播机制下,被害人难以在海量的网络暴力信息中确定每一个账号的评论、点赞、转发在整个网络暴力事件所产生的具体作用,无力调取被平台限流、屏蔽、删除的网络暴力信息,更无法找到网络暴力信息传播链条中的“零号施暴者”。平台特殊的媒介环境决定了被害人搜集证据往往需要付出较大的经济成本和社会成本,“造谣一张

嘴、辟谣跑断腿”堪称是当事人自力救济困难的准确白描,被害人在平台上发布维权声明的行为甚至会招致“玻璃心、博关注、气急败坏”等负面评价。在回应无力且救济乏术的情况下,被害人在面对网络暴力信息聚集而成的舆论压制时,经常陷入“孤立无援”的绝望心理,将网络暴力的社会戾气归咎为自身的个人问题,导致网络暴力的“二次伤害”。例如2022年发生的粉色头发女孩被网暴案中,面对网民将“染粉色头发”的女孩污蔑为“从事不正经职业”“陪酒女”的言论攻击,被害人的维权措施尽管已经“倾其所有”,但在平台海量的网络暴力信息中,仅仅能够找到少数施暴者,根本无力扭转网民的偏见与污名,被害人在平台中发布维权声明、评论的做法甚至被嘲讽为“玻璃心、想不开”。最终,压倒性、不可逆的长期舆论制裁诱发了被害人的自杀后果。

三、引入媒介正义:网络暴力信息治理义务的内涵重塑

基于平台在网络暴力信息生成、扩散、异化过程中存在的媒介不公现象,平台承担网络暴力信息治理义务的制度意义就不应仅仅是宏观层面的看门人治理,而是要通过平台媒介角色的重塑,搭建优化内容质量、鼓励理性讨论、提供弱者救济的媒介正义环境,避免网络暴力信息通过平台的媒介化作用机制对个体和社会造成伤害,以使平台这一媒介型基础设施能够真正服务于网络信息内容的生态优化和数字时代人类福祉的增益。由此,平台履行网络暴力信息治理义务就具备了媒介正义的理论基础和价值意蕴。在网络暴力信息生成、扩散、异化的三个核心环节,媒介正义提出了旨在矫正媒介不公的平台治理要求,为网络暴力信息治理义务注入了全新的理论内涵^[35]。

(一) 基于媒介信息正义的网暴内容针对性审查

在网络暴力信息的生成环节,《征求意见稿》第三章设置了以网络暴力信息统计、涉网络暴力

重点账号管理、网络暴力信息预警为代表的监测预警义务。面对平台鼓励偏激性内容生产的媒介不公问题，媒介信息正义要求平台承担与内容生产风险相匹配的内容质量管理责任，为网络暴力信息的监测预警义务注入了针对性审查的独特内涵。

媒介信息正义的提出，旨在矫正平台鼓励内容生产数量、忽视内容生产质量的媒介环境，对平台提出了实现内容生产激励与内容生态健康均衡发展的全新要求。媒介信息正义强调网络信息内容的品质与安全^[36]，平台作为数字媒介，虽然可以自主管理、引导用户形成内容创作的积极性，以便在“内容为王”的数字经济环境下打造内容营销的主体优势，实现用户内容生产的可持续增长^[37]，但不可以用鼓励偏激性言论生产、损害网络信息内容品质的方式，在内容资本积累的过程中纵容语言暴力的横行^[38]。因此，平台需要承担与鼓励内容自由创作相匹配的内容质量监督责任：平台的用户管理机制营造了何种网络信息内容生产的主体行动策略，就需要承担与这一内容生产秩序相适配的媒介公共治理责任；平台的话题组织机制塑造了何种内容特色，就需要充分监测与这一内容特色相伴相生的网络暴力风险。

媒介信息正义要求平台承担与内容生产风险相匹配的内容质量管理责任，为网络暴力信息的监测预警义务注入了针对性审查的独特内涵。所谓针对性审查，即平台需要结合自身内容业务运营的特点，归纳、总结平台内部容易出现的网络暴力信息类型，设置针对性的网络暴力信息监测预警机制，以及时、准确地识别平台内容业务存在的网络暴力风险。长期以来，我国的新型网络法为平台配置的主要是一般性的违法犯罪内容审核义务，全面审核是平台履行此类义务的基本要求^[39]。扩大对电信网络诈骗、网络赌博、色情暴力等违法或不良信息的审核覆盖面，成为平台内容合规、履行犯罪预防义务的通用做法。然而，在媒介信息正义的要求下，监测预警义务并不意味着面面俱到地审查海量的网络暴力信息，而是针对性地考虑平台的内容业务特征、用户群

体特点、社区讨论模式等因素，完善与平台内容业务具有高度关联性的网络暴力信息库和案例样本，排查、处理经常在平台中发布网络暴力信息的重点账号，精准识别与平台内容业务高度相关的网络暴力风险，使监测预警义务的履行能够与平台内容业务的特殊性相适配。由此，针对性审查的义务就在平台市场化的内容运营逻辑与公共性的内容质量目标之间建立起直接联系，指向了平台业务经营与社会公共责任均衡发展的新目标^[40]。

（二）基于媒介程序正义的网暴舆情去极化处置

在网络暴力信息的扩散环节，《征求意见稿》第四章规定了以网络暴力信息应急处置、网络暴力话题管理、直播和短视频审核为代表的平台事中处置义务。面对平台流量化运营过程中极化情绪快速扩散的问题，媒介程序正义要求平台优化网络信息内容的传播机制，防止极端化的网络暴力舆情在争议事件中快速传播。

媒介正义中强调信息公正传播过程的程序正义，旨在通过平台对网络信息内容传播机制的自我管理和动态优化，矫正平台诱导网络暴力信息快速扩散的媒介不公角色，引导网民在平台所搭建的“正义框架”下展开争议事件的讨论。媒介程序正义要求媒介的信息传播过程尽可能客观、公正，不可滥用媒介权力诱导公众形成有违社会正义的认知和理解^[41]。在流量逻辑的驱使下，平台实质上放大了网络暴力信息的扩散力、影响力，受众的注意力在平台流量化的媒介信息传递逻辑下集中在极端化情绪的共情和宣泄，忽视了对争议事件本身的理性思考^[42]。媒介程序正义正是看到了平台对流量的过度追逐和对媒介权力的滥用，主张通过平台内容治理义务的施加，引导平台自我矫正争议事件的网络信息内容传播机制，深度打捞那些淹没在极端化舆情言论之中的理性思考，引导受众发现争议事件本应具有的多元、复杂面向，形成充满正义性、富有建设性的争议事件讨论过程。

媒介程序正义对于争议事件讨论过程的公正性强调，为平台对涉网络暴力舆情的事中处置

提出了去极化的义务履行目标。所谓去极化处置,即平台对于网络暴力信息、话题等内容的管理和处置应以有效控制涉网络暴力舆情的极端化发展为目标,要求平台具备健康引导争议事件讨论议程的能力。在网络违法犯罪的治理中,平台常规的内容处置义务主要以对抗性的屏蔽、删除措施为核心手段,意在及时消除用户违法违规行为所产生的影响^[43],释放出必要的秩序控制功能。如果说传统内容处置义务的内涵在于“控制”,那么网络暴力信息治理义务的平台角色则在于“建构”。去极化处置的义务内涵明显指向了平台对争议事件讨论秩序的优化和引领,平台的义务履行不仅满足于对网络暴力信息消极、被动的屏蔽删除,还在于主动、直面一个相对理性、观点多元的争议事件讨论过程。在正向的制度构建目标下,平台需要以事中处置义务的履行落实对热搜、话题、社区、群组等内容场景的纵深管理,自我矫正过度流量化的网络信息内容传播机制。唯有如此,平台才能深度改变网络暴力信息的优势传播地位,让理性化的网络信息内容在健康的讨论秩序中重新回归大众的视野。

(三) 基于媒介分配正义的网暴对象倾斜式保护

在网络暴力的事后救济环节,《征求意见稿》第五章系统设置了网络暴力信息防护、网络暴力私信管理、网络暴力投诉举报响应、取证协助、特殊群体保护等一系列极富特色的被害保护义务。由于平台的媒介呈现机制在实质上造成加害人对被害人单方面的数字权力宰制,媒介分配正义要求平台面向被害人提供倾斜性的保护机制和治理资源,为被害人保留最低限度的数字化生存空间。

媒介分配正义强调重新分配双方的媒介话语权,通过媒介权力的重组实现对数字弱势群体的特殊关怀^[44]。媒介分配正义要求平台对被害人提供倾斜式保护,为被害人调配必要的媒介资源,协助被害人对抗加害人的数字临时性权力,保护被害人数字化生存所需要的最低限度空间,以期通过差异化的媒介资源分配矫正失衡化的媒介权力赋能。面对加害人通过平台的评论、点

赞、转发等媒介通道形成的数字临时性权力,平台倾斜式保护旨在为被害人调配必要的媒介资源,使被害人的发声、被害人的行动获得应有的重视,瓦解加害人形成情感同盟和舆论审判的媒介权力基础;提升被害人寻求媒介对话、展开权利救济的可能性,震慑、对抗网络暴力信息之中由加害人掌握的数字临时性权力,使双方的生活处境、力量对比回到基本对称的状态^[45]。更为重要的是,平台的主动进场能够对私信轰炸、人肉搜索等权力宰制形式形成必要的技术反制,为被害人保留最低限度的数字化生存空间,防止被害人被网络暴力信息群聚而成的舆论道德审判压垮^[46]。从数字权利的角度来看,平台的被害人保护义务实际上与数字时代愈发重要的数字化生存联系在一起,彰显出数字文明的时代价值。

媒介分配正义对网络暴力被害人这一特殊数字弱势群体的情景关怀,为网络暴力信息治理义务注入了倾斜式保护的制度内涵。所谓对网暴对象的倾斜式保护,即为网络暴力被害人提供更加主动、有效乃至家长主义的权益保护机制。在一般性的网络犯罪治理义务中,平台对网络违法犯罪活动的响应处于相对被动的状态,平台义务的履行多以监管部门或执法部门发送网络违法犯罪活动的控制指令为前提。与被动的响应控制不同,倾斜式保护意味着平台要主动协助网络暴力被害人的权利救济活动,实际上为平台提出了更高的注意义务要求。一方面,平台需要专门搭建防止被害人遭受网络暴力信息侵害的技术预防机制,在用户界面中嵌入网络暴力被害预防的应用功能,打造类似于预防青少年网络欺凌现象的“反欺凌计划”(anti-bullying programs)^[47],为被害人的权利救济提供必要的平台制度支撑和媒介资源支持。另一方面,平台对网络暴力信息的响应不再被动等待监管部门或执法部门的控制指令,而是更加积极、主动地回应被害人对平台提出的权利救济诉求。对于未成年人、老年人作为网络暴力被害人的特殊情形,平台保护义务甚至需要具备强烈的“法律家长主义”色彩,主动提示、积极协助此类特殊被害人群体开启网络暴力信息防护机制。总之,倾斜性保护的义务内

涵意味着平台不再是作为单纯负责内容呈现和内容转发的中立技术媒介^[48]，而是通过被害预防的义务履行将媒介资源向被害人倾斜，全力扭转网络暴力冲突中加害人与被害人权力不对等的非均衡态势。

四、整合媒介权力：网络暴力信息治理义务的法治实现

媒介正义虽然为网络暴力信息治理义务注入了全新的理论内涵和履行目标，但平台如何履行此类义务、如何落实主体责任，尚缺乏具体的技术路线。由于以《征求意见稿》为代表的法律具备显性的媒介公共权力，掌握媒介技术优势和资源优势的平台具备隐性的媒介技术权力^[49]，平台对网络暴力信息治理义务的有效履行就取决于公共权力和技术权力两种不同媒介权力的有效整合。这也为平台治理勾勒出新的技术路线——法律通过外部媒介公共权力的合理施加，引导平台合理运用内部的媒介技术权力优势，自主形成网络暴力内容治理的合规体系，在国家管控、平台自治的深度沟通和有效联动中实现复杂的规制目标^[50]。实际上，《征求意见稿》第二章已经从建立完善网络暴力信息治理机制、强化网络暴力账号信息管理、完善网络暴力平台管理公约、定期发布网络暴力信息治理报告等角度，设置了以平台内容合规为核心的间接规制框架，体现出整合媒介权力的制度意蕴。

（一）事前预警义务：平台内容生态的针对性监测

由于不同平台的业务靶向、用户结构、内容数据、分析模型都存在显著差异，网络暴力信息治理义务既难以要求平台形成统一固定的网络暴力信息监测预警标准，又无法直接判断平台某种监测预警机制是否绝对合法或非法，因此，较为务实的规制策略是建立类似于产品质量监测的义务审核方式，对平台监测预警的针对性进行质量评级，通过合理的外部规制压力引导平台主动调动自身的媒介技术资源，从而完善网络暴力信息监测预警的内容合规管理体系。

第一，平台的内容审核机制与自身业务经营的内容靶向是否适配。尽管目前各大平台都设置了由机器审核和人工巡查组成的内容安全审核机制，但这些数据库的内容样本主要服务于平台整体的内容质量审核，难以针对性识别、监测平台内部的网络暴力信息风险。因此，平台的网络暴力信息监测预警能否针对性识别平台内部独特的网络暴力风险，成为内容审核环节合规体系建设的关键^[51]：其一，平台是否设置与内容业务具有明确关联性的网络暴力信息审核机制。例如，B站针对自身内容业务存在的“涉性”网络暴力风险开展了女性话题治理专项行动，由风纪委员会判断弹幕、评论是否存在侮辱女性的情感倾向，即从人工审核的角度强化了网络暴力信息审核的内容针对性。其二，平台是否对自身网络暴力的内容审核机制进行巡回审查和内容更新。由于网络暴力的攻击形式随着时间的推移、内容主题的更新而变化^[52]，平台迫切需要对内容审核机制进行必要的巡回审查、定期更新内容审核规则、随机捞回必要的网络暴力信息，以审核内容模型的准确性和针对性。其三，平台是否强化对高风险内容运营场景的网络暴力信息审核力度，即对于热搜话题、直播PK、短视频新闻等容易引发网络暴力冲突的内容场景，平台是否设置了特殊的技术审和人工审规则，是否在内容合规管理体系中配备专门的内容审核机制。

第二，平台的重点用户账号管理是否与用户的内容创作门槛相适应。平台用户内容创作、发布信息门槛的高低，通常会塑造不同健康程度的用户内容创作生态。因此，平台账号管理的合规标准应以用户内容创作门槛为核心，根据平台不同的内容创作环境配置不同的用户管理义务要求。对于用户内容创作自由度较高的平台，其账号管理重点在于注册成本和发言成本较低的新号、僵尸号，防止此类用户、账号在平台肆意发布网络暴力信息。例如，小红书根据账号的使用频率、浏览时间等维度筛出僵尸号或水军炒作号，并对上述账号发布的笔记限制曝光量；TikTok也采取技术审核手段实时审核机器人攻击、刷流量、恶意评论等重点账号，并采取限流、限制访

问平台部分功能等账号管理措施。对于用户内容创作门槛较高的平台,由于用户往往需要进行严格的账号认证、达到特定的账号级别、通过严苛的答题测试才能进行内容创作,出现新号、僵尸号发布网络暴力信息的可能性较低,故此类平台无须承担过重的用户账号管理义务,仅需在特定领域、特定话题中动态监测、管理具有较高发言权的意见领袖即可。

(二) 事中处置义务: 极端网暴舆情的有效性引导

对极端网络暴力舆情的议程设置引导,指平台需要搭建有利于争议事件正常化、多元化探讨的议程框架和讨论秩序,以事中处置义务的履行有效抑制网络暴力信息的快速扩散。由于不同平台在内容运营、话题群组、社区氛围方面具有较大的差异性,事中处置义务不宜提出整齐划一的义务履行要求,而是设置总体性的议程设置框架,引导平台在这一框架下结合自身的内容业务场景配备具体的网络暴力信息处置措施。

第一,在表层的网暴舆情引导环节,平台应当建立以“降温”为核心的重大网络暴力舆情的应急处置机制,避免极端化的网络暴力信息主导争议事件的讨论议程。事中处置义务要求平台建立专门的涉网络暴力舆情处置板块,引导平台以网络暴力信息指向的争议事件冲突和关键人物为审查重心,发现网络暴力信息得以快速扩散、全网传播的关键舆情节点,并对关键舆情节点的相关网络暴力信息及时打标、分类分级,采取对应的删除、限制传播、搜索降权等“降温”措施,从而通过平台的媒介作用力深度影响争议事件的网络舆情走向,释放平台对于网络信息内容生态的秩序建构功能^[53]。至于网络暴力的分类分级种类、对应处置机制、舆情干预策略、团队人员职责、内部操作规范等事项,事中处置义务应当允许平台自主探索更加灵活、多样的形式和方法^[54],网信等监管部门不可过度干涉。

第二,在深层的关联话题管理环节,平台应当建立以“脱钩”为导向的网络暴力舆情关联话题管理机制,防止网络暴力舆情与常见的争议话题产生二次关联。“脱钩”的话题管理机制要求

平台分析网络暴力舆情中的社会舆论基础,及时发现网络暴力信息中潜在的偏见、对立话题。至于“脱钩”具体措施,平台可探索网络暴力风险浮动提示、恶意话题实时熔断、意见领袖言论重点监测、对关联话题热度降级等多元化的内容合规方案。当然,鉴于“脱钩”处置可能会对用户的内容创作造成明显干预,平台需要完善具有实质交互性的申诉救济机制,优化关联话题内容先审后发、意见领袖发言自律等治理措施^[55],避免平台因过度履行事中处置义务而损伤正常的网络信息内容生态。

第三,在源头的网络暴力争议事件澄清环节,平台应当建立以“真相”为核心的争议事件讨论议程,以多元的信息披露回应社会激增的信息需求。在争议事件多维度讨论的议程设置框架下,平台需要结合自身话题讨论、内容生产的特点,丰富争议事件的信息披露视角、延长争议事件缘起发展的时间轴:如果争议事件呈现为双方冲突但只有单方回应,平台需要在议程设置中设置争议标签,显著标注“另一方尚未回应”,以削弱单方披露对于相关受众的情绪渲染能力;如果争议事件体现为多阶段、多流程的发展走向,平台应该在信息发布板块上线争议事件的时间轴和流程线等功能板块,清晰标注、显著呈现争议事件的多元信息维度;如果争议事件由自媒体发布,平台应当要求自媒体准确标注信息来源,并在用户界面显著体现自媒体标识,以强调对争议事件信息源的真实性管理。

(三) 被害保护义务: 平台在线预防的及时性响应

平台的被害保护义务应强调被害人救济优先。一旦网络暴力信息的聚众现象在平台中形成规模,当事人将面临数字生命被剥夺乃至身心遭受严重损害的特殊状态。因此,平台必须第一时间采取应急保护措施,最大限度地保护被害人免受网络暴力信息聚集的大规模侵害。并且,在被害保护环节提高义务的履行标准、提升网信部门的监管强度、增加平台事后治理的内容合规成本,还在于通过外部的媒介公共权力倒逼平台将更多的合规资源投入网络暴力信息的事前、事中

预防环节，实现预防治理的现代化治理目标。

第一，平台对被害人的信息保护和私信管理应满足“屏蔽+过滤”的双线防护标准。所谓“屏蔽”，即平台私信规则的设置应满足网络暴力被害人仅接收好友私信或拒绝接收所有私信的基础功能，能够为被害人保留最低限度的数字化生存空间。2022年3月，国内各大社交媒体平台陆续上线了私信屏蔽功能，其中小红书的私信限制功能具有较高的参考价值，小红书的用户如果收到陌生人私信且并未回复，则陌生人无法继续向该用户发送私信，并且陌生人已经发送的私信也不会直接呈现在用户界面。所谓“过滤”，即设置具有自动检测能力的网络暴力技术反制措施^[56]，有效过滤私信、评论等环节的网络暴力信息。例如，抖音针对实施用户的语言暴力行为和不友善行为上线了“发文警示”“负向内容过滤”等8项内容治理措施，新浪微博、Instagram等平台同样采取智能化的违禁词识别，提示用户是否发布攻击性言论。随着内容反制技术的日趋成熟，被害预防反而将越来越多地从施暴者角度进行，基于过滤的主动技术反制措施将在网络暴力信息治理中发挥更大的作用。

第二，举报响应制度应强调举报的便利性和受理的及时性。举报的便利性，即平台应该根据网络暴力超大规模侵权的特质，为用户提供网络暴力信息批量举报的窗口，降低举报维权的难度。近年来，诸多平台上线并完善了批量举报投诉功能：豆瓣上线的一键举证功能，可以支持用户批量多选私信聊天记录并向平台举报；抖音则允许用户自主设置网络暴力信息生成、传播的起始时间，允许用户对特定时间段的批量私信攻击行为一键举报。而对举报内容的响应，平台需要根据自身的内容审核能力，在特定的时限内对举报内容进行反馈和处置。考虑到被害人保护的迫切需要，平台应该在24至48小时内回复被害人。若举报内容属实，平台应立即采取相应的处置措施。并且，按照2023年9月网信办发布的《关于进一步加强网络侵权信息举报工作的指导意见》要求，平台对时效性强、举证时间久、可能给举报人造成较大负面影响的侵权信息，还应先

行采取“加私”措施及时阻断相关信息分享传播。作为被害人向平台寻求权利救济的重要制度通道，平台拒绝举报响应的情形十分有限，只有当用户多次以无明显根据的样本举报之时，平台才可根据“受信任举报人原则”减少回复、审核的频率^[57]。

第三，协助取证义务中平台的技术协作范围应覆盖网络暴力信息的内容和传播两个核心环节。网络暴力信息的“一键取证”等技术协助的范围不仅限于提供平台现存、公开的网络暴力信息，还需将平台删除、屏蔽等被害人难以收集的网络暴力信息一并提供。此外，对于网络暴力信息复杂的舆情传播过程、网络暴力信息关键的传播节点、网络暴力信息的舆情图谱等内容，平台仍需协助被害人的证据调取。以被害人为中心的协助取证义务标准不仅能在网络暴力救济的“黄金时段”给予被害人积极的人格权救济反馈、对网络暴力信息的无序生产予以必要的震慑，还可以进一步审查平台对网络暴力信息的处置规范情况，通过外部的监管压力与规制压力敦促平台在事前、事中有效落实主体责任。

在法律责任层面，随着《征求意见稿》的出台，平台未能有效形成基于网络暴力信息治理的事前预警、事中处置、事后保护体系，将承担明确的法律后果：平台未履行网络暴力信息治理义务、给他人造成损害的，平台应对其不作为造成的损害扩大部分承担间接侵权责任；平台日常性的内容合规体系未达到网络暴力信息治理义务履行标准的，应承担《征求意见稿》第二十六条规定的行政法律责任；平台基于蹭炒热度、推广引流等目的，不依法履行网络暴力信息治理义务、经监管部门责令采取改正措施而拒不改正，并导致网络暴力信息大量传播的，还应承担以拒不履行信息网络安全管理义务为核心的刑事法律责任。

五、结语

《征求意见稿》中对网络暴力信息治理义务的增设，标志着数字时代的网络暴力规制与网络

信息内容在平台媒介中的数字化传播过程产生了制度联系,网络暴力的法律规制呈现出多领域、交叉性的全新特点^[58]。面对法律规制转型与网络内容治理的牵连交织,网络暴力信息治理义务的分析如果仅依靠法学研究的内部逻辑推演,恐怕难以在高度复杂的内容传播机制中找到平台义务履行的真正发力点。正视、尊重平台内容运营的客观规律和独特机理^[59],已然成为数字时代网络暴力研究的逻辑前提。基于此,数字媒介理论这一曾经与法学研究联系极度微弱的传播学话语,经由网络暴力信息治理义务的问题意识牵引,为数字时代法学研究的转型提供了交叉学科的独特视野。在法学与传播学的深度对话和彼此审视中,网络暴力信息激增这样一个横跨多个领域、牵扯多方主体的复杂现象,能够在平台的媒介不公视角得到新的理论解释;网络暴力信息治理中法律正义话语与媒介伦理观念的盘桓交织,得以在媒介正义的概念框架下进行新的知识重组;网络暴力信息治理义务中平台如何自我优化媒介环境、重塑媒介交往秩序的治理难题,有望在整合媒介权力的规制技术中贡献新的观点。在“内容为王”的数字时代,“法学+传播学”的理论潜力或许不止于此,随着中国近年来网络信息内容治理的法治化进程加速,重大公共事件讨论的意见极化、自媒体内容创作的言论边界、网络粉丝的饭圈治理、生成式人工智能的内容规制等一系列涉及数字传播规律的法治问题,均需在法学与传播学的交叉地带探索更加数字化、更具想象力的答案。

参考文献:

- [1] 单勇. 数字平台与犯罪治理转型[J]. 社会学研究, 2022, 37(4): 45-68, 227.
- [2] 石经海. 论网络暴力的实质与刑法适用规则的完善[J]. 法律科学(西北政法大学学报), 2023, 42(5): 69-82.
- [3] 车浩. 诽谤罪的法益构造与诉讼机制[J]. 中国刑事法杂志, 2021(1): 63-77.
- [4] 于冲. 网络“聚量性”侮辱诽谤行为的刑法评价[J]. 中国法律评论, 2023(3): 87-98.
- [5] 徐颖. 论“网络暴力”致人自杀死亡的刑事责任[J]. 政法论坛, 2020, 38(1): 132-142.
- [6] 陈罗兰. 网络暴力的刑法治理与罪名增设[J]. 法律科学(西北政法大学学报), 2023, 42(5): 115-123.
- [7] 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.
- [8] 江溯. 论网络犯罪治理的公私合作模式[J]. 政治与法律, 2020(8): 38-52.
- [9] 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.
- [10] CIOFFI J W, KENNEY M F, ZYSMAN J. Platform power and regulatory politics: Polanyi for the twenty-first century[J]. New Political Economy, 2022, 27(5): 820-836.
- [11] 敬力嘉. 网络暴力公私合作治理模式的反思与调试[J]. 江汉论坛, 2023(5): 136-144.
- [12] 刘艳红. 理念、逻辑与路径: 网络暴力法治化治理研究[J]. 江淮论坛, 2022(6): 21-30.
- [13] 邵登辉. 群体性网络暴力治理——以网络平台风险预防义务为视角[J]. 中南民族大学学报(人文社会科学版), 2023, 43(7): 97-105, 185.
- [14] PLANTIN J C, PUNATHAMBEKAR A. Digital media infrastructures: Pipes, platforms, and politics[J]. Media, Culture & Society, 2019, 41(2): 163-174.
- [15] LOGAN R K. Understanding humans: The extensions of digital media, information [EB/OL]. (2019-09-29) [2023-09-20]. <https://doi.org/10.3390/info10100304>.
- [16] MADIANOU M, MILLER D T. Polymedia: Towards a new theory of digital media in interpersonal communication[J]. International Journal of Cultural Studies, 2013, 16(2): 169-187.
- [17] 孟超. 网络内容规范中的隐喻和想象: 对政府和互联网平台的话语分析[J]. 新闻记者, 2023(6): 46-56.
- [18] 卡罗琳·艾姆克. 何故为敌[M]. 郭力, 译. 北京: 社会科学文献出版社, 2015: 5.
- [19] 尼克·库尔德利, 安德烈亚斯·赫普. 现实的中介化建构[M]. 刘泱育, 译. 上海: 复旦大学出版社, 2023: 274.
- [20] BOSSETTA M. The digital architectures of social media: Comparing political campaigning on facebook, twitter, instagram, and snapchat in the 2016 U.S. election[J]. Journalism & Mass Communication Quarterly, 2018, 95(2): 471-496.
- [21] 韩炳哲. 在群中: 数字媒体时代的大众心理学[M]. 程巍, 译. 北京: 中信出版集团股份有限公司, 2019: 3-9.
- [22] 刘少杰. 网络社会的结构变迁与演化趋势[M]. 北京: 中国人民大学出版社, 2019: 91.

- [23] 南希·K·拜厄姆. 交往在云端：数字时代的人际关系[M]. 董晨宇, 唐悦哲, 译. 北京：中国人民大学出版社, 2020: 62.
- [24] 尚帅. 传播的兴趣偏向和浑然不觉的符号暴力——基于《今日头条》的算法控制研究[J]. 现代传播(中国传媒大学学报), 2018, 40(10): 58-62.
- [25] 解庆锋. 网民信念沟：媒介接触对网民意见分歧的影响[J]. 新闻与传播研究, 2022, 29(7): 55-74, 127.
- [26] ELLIS J R. More than a trivial pursuit: Public order policing narratives and the ‘social media test’[J]. *Crime Media Culture*, 2021, 17(2): 185-207.
- [27] 陈启涵. 平台可供性视角下网络对立情绪的流量政治[J]. 新闻界, 2023(1): 78-87, 96.
- [28] POWELL A, STRATTON G, CAMERON R. Digital criminology: Crime and justice in digital society[M]. New York: Routledge, 2018: 112-118.
- [29] 刘绩宏. 网络谣言到网络暴力的演化机制研究[J]. 当代传播, 2016(3): 83-85.
- [30] 药琦, 谢紫怡. “心惊暴”迷思：从刘学州事件看世俗媒体批评[J]. 新闻记者, 2023(4): 3-13, 96.
- [31] 罗譔. 网络暴力的微观权力结构与个体的“数字性死亡”[J]. 现代传播(中国传媒大学学报), 2020, 42(6): 151-157.
- [32] 晏齐宏. 新媒体表达中的情感、框架及注意力：基于计算传播路径的探索[M]. 上海：复旦大学出版社, 2022: 186.
- [33] 冯济海. “粉”“黑”之争：网络迷群极化的“日常化”转向[J]. 社会学研究, 2021, 36(6): 113-135, 228.
- [34] WILLIAMS M L, BURNAP P, JAVED A, et al. Hate in the machine: Anti-black and anti-muslim social media posts as predictors of offline racially and religiously aggravated crime[J]. *British Journal of Criminology*, 2020, 60(1): 93-117.
- [35] 齐爱军. 尼克·库尔德利“媒介正义”观评析[J]. 新闻与传播研究, 2021, 8(3): 39-56, 126.
- [36] 袁靖华. 论媒介正义的概念及其维度——基于拉斯韦尔“5W”传播模式[J]. 国际新闻界, 2012, 34(4): 34-39.
- [37] HASINOFF A A, SCHNEIDER N. From scalability to subsidiarity in addressing online harm [EB/OL]. (2022-09-01) [2023-09-20]. <https://doi.org/10.1177/20563051221126041>.
- [38] FUCHS C. Foundations of communication/media/digital (in)justice[J]. *Journal of Media Ethics*, 2021, 36(4): 186-201.
- [39] 单勇. 论互联网平台的犯罪控制义务[J]. 现代法学, 2022, 44(3): 66-81.
- [40] 刘权. 论互联网平台的主体责任[J]. 华东政法大学学报, 2022, 25(5): 79-93.
- [41] 尼克·库尔德利. 媒介、社会与世界：社会理论与数字媒介实践[M]. 何道宽, 译. 上海：复旦大学出版社, 2022: 197-202.
- [42] 玛丽·吉科. 超链接：互联网、数字媒体和技术——社会生活[M]. 黄雅兰, 译. 北京：清华大学出版社, 2019: 142-143.
- [43] 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.
- [44] 袁靖华. 媒介正义论：走向正义的传播理论与实践[J]. 国际新闻界, 2011, 33(2): 25-30.
- [45] 约翰·罗尔斯. 正义论[M]. 何怀宏, 何包钢, 廖申白, 译. 北京：中国社会科学出版社, 2009: 10.
- [46] 伍德志. 网络社会道德的普泛化及其法律规制[J]. 法商研究, 2023, 40(4): 72-86.
- [47] HAY C, MELDRUM R. Bullying victimization and adolescent self-harm: Testing hypotheses from general strain theory[J]. *Journal of Youth and Adolescence*, 2010, 39(5): 446-459.
- [48] BIJLENGA N, KLEEMANS E R. Criminals seeking ICT-expertise: An exploratory study of Dutch cases[J]. *European Journal on Criminal Policy and Research*, 2018, 24(3): 253-268.
- [49] 谢登科. 网络暴力犯罪的公私协同治理模式[J]. 法律科学(西北政法大学学报), 2022, 40(2): 92-101.
- [50] 科林·斯科特. 规制、治理与法律：前沿问题研究[M]. 安永康, 译. 北京：清华大学出版社, 2018: 122.
- [51] MIRO-LLINARES F, MONEVA A, ESTEVE M. Hate is in the air! But where? Introducing an algorithm to detect hate speech in digital microenvironments [EB/OL]. (2018-11-15) [2023-09-20]. <https://doi.org/10.1186/s40163-018-0089-1>.
- [52] PATTON D U, HONG J S, RANNEY M, et al. Social media as a vector for youth violence: A review of the literature[J]. *Computers in Human Behavior*, 2014, 35: 548-553.
- [53] 邹晓玫. 网络服务提供者之角色构造研究[J]. 中南大学学报(社会科学版), 2017, 23(3): 63-69.
- [54] 朱笑延. 数字治理的刑法介入模式——以电信网络诈骗的治理变革为例[J]. 华中科技大学学报(社会科学版), 2023, 37(2): 30-40.
- [55] 赵精武. 异化的网络评论——再论网络暴力信息的阶

- 段化治理[J]. 北方法学, 2023, 17(5): 21–36.
- [56] KOTSAKIS R, VRYSIS L, VRYZAS N. A web framework for information aggregation and management of multilingual hate speech [EB/OL]. (2023-06-29) [2023-09-20]. <https://doi.org/10.1016/j.heliyon.2023.e16084>.
- [57] QUINTAIS J P, SCHWEMER S F. The interplay between the digital services act and sector regulation: How special is copyright[J]. *European Journal of Risk Regulation*, 2022, 13(2): 191–217.
- [58] 刘艳红. 网络暴力治理的法治化转型及立法体系建构[J]. *法学研究*, 2023, 45(5): 79–95.
- [59] 张新平. 以平台为重点: 网络社会法律治理的新思路[J]. *中南大学学报(社会科学版)*, 2022, 28(2): 116–123.

Towards media justice: Platform role and law implementation of cyber violence information governance

ZHU Xiaoyan

(School of Law, Jilin University, Changchun 130012, China)

Abstract: The promulgation of the obligation of cyber violence information governance marks a "super responsible entity" for the platform. Understanding the internal regularities of generation, diffusion, and alienation of cyber violence information under platform environment has become the key to understanding the regulation transformation of cyber violence in the Digital Age. By introducing the digital media theory, it can be found that cyber violence information is not spontaneously gathered by "the Crowd", but grows in a media environment where platforms encourage extremist speech, support radical emotions, and promote public opinion clusters. To build a media justice order that optimizes content quality, encourages rational discussion, and provides victim relief has become the unique connotation that the platform implements its duty in governing cyber violence. Before the incident, the platform's early warning system should aim to configure a content supervision mechanism applicable to its content ecology, to perceive the risk of cyber violence information at real-time; in the process of the incident, an agenda framework should be set up to healthily guide public opinion orientations and to effectively block the spread of online violence information; and at the stage of protecting the victim, technical counter-measures with timely response ability should be taken to prevent victims from public bullying.

Key Words: cyberviolence information; platform obligations; media justice; content compliance

[编辑: 苏慧]