


社会问题研究之 网络暴力

2024年6月



网络暴力的新特点

- 商业利益驱动下的大V、网红推动
- 网暴黑色产业链形成
- 集体狂欢下的网络暴力
- 连环暴力与叠加暴力
- 维权空白
- 网络暴力工具化、日常化

网络暴力的心理机制

- 群体极化是“团体成员一开始即有某些偏向，在商议后，人们朝偏向的方向继续移动，最后形成极端的观点”。（桑斯坦）
- 网络去抑制效应：勒庞提出，一个群体的成员身份会导致主体处于“淹没状态”，在这种状态下，对个体行为的一般约束消失了。网络的匿名性和广场效应也具有削弱对个体行为约束的“去抑制”效果。
- 心理宣泄：也称作“霍桑效应”。工作的过程就是积累压力的过程，每个人的心理承受能力都是有限的，若不及时排解压力，就会被压垮。

网络暴力治理的法律体系

- 《网络安全法》是网络暴力治理的指引性法律，主要用以明确国家在该领域的方针、政策与治理原则；
- 《民法典》《刑法》《治安管理处罚法》及相关司法解释，是执法和司法领域中解决纠纷的主要依据，用以明确和分配特定主体的责任、权利。
- 《全国人民代表大会常务委员会关于维护互联网安全的决定》《互联网信息服务管理办法》《互联网直播服务管理规定》《互联网视听节目服务管理规定》《互联网用户账号名称管理规定》《网络信息内容生态治理规定》等是维护网络平台日常运行的规范性文件。



事后救济：当前网络暴力的法律规制特点

- 事后救济原则
- 遵循“不告不理”原则
- 遵循个体主义治理模式（自诉为主）
- 个案救济为主
- 强调事后追责与赔偿
- 以救济被害人权利为重心

网络暴力的法律标准逐渐明晰

“网络暴力针对个人集中发布侮辱谩骂、造谣诽谤、侵犯隐私等违法信息及其他不友善信息”。

（《网暴治理通知》）

“在网络上针对个人肆意发布谩骂侮辱、造谣诽谤、侵犯隐私等信息的网络暴力行为，贬损他人人格，损害他人名誉”。（“惩治网暴20条”）

“通过网络对个人集中发布的，侮辱谩骂、造谣诽谤、侵犯隐私，以及严重影响身心健康的道德绑架、贬低歧视、恶意揣测等违法和不良信息。”（《规定》）



网暴程度？

网暴形式？

网暴后果？

网上暴力与

网下暴力的

关系？

网络暴力的规制

- 2022年11月，中央网信办出台《网络暴力治理通知》
- 《网络暴力信息治理规定》
- 国家网信办于2022年11月出台了《关于切实加强网络暴力治理的通知》
- 2023年以来，公、检、法三部门联合出台了《关于依法惩治网络暴力犯罪的指导意见》

网络暴力立案标准

- 点击、浏览或转发次数达到一定标准：同一诽谤信息实际被点击、浏览次数达到五千次以上，或者被转发次数达到五百次以上的。
- 造成严重后果：造成被害人或者其近亲属精神失常、自残、自杀等严重后果的。
- 二年内曾因诽谤受过行政处罚，又诽谤他人的。
- 其他情节严重的情形。

案例：网络诽谤自诉转公诉第一案

- ▶ 杭州女子取快递被造谣案是发生于2020年7月中华人民共和国浙江省杭州市余杭区的一起网络诽谤案，为2020年度十大法律监督案例之一，也是该国首起司法机关针对普通公民的网络诽谤提起公诉的案件。



明确“自诉转公诉”的程序衔接问题；明确“自诉转公诉”的公诉标准问题；畅通被害人的权利救济途径。

互联网平台：网络暴力治理的“守门人”

- 平台技术治理成为目前的主要方案并被相关立法所规定。技术治理可实现对网络暴力的事前预防及信息准确识别。
- 立法规定了网络暴力识别模型、用户保障机制、信息阻断等。其中，建立网络暴力识别模型是关键，其包括对内容的识别模型和对高危账号的识别模型。

“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，综合考虑事件类别、针对主体、参与人数、信息内容、发布频次、环节场景、举报投诉等维度，及时发现预警网络暴力风险。”（《网络暴力信息治理规定》（征求意见稿））

网络中立：ISP的无差别对待原则


- “网络中立原则”是2003年由美国哥伦比亚大学媒体法律学教授吴修铭（Tim Wu）提出的，即互联网服务提供商必须同等对待来自各方的所有内容。
- 2014年奥巴马政府为FCC建议的**网络中立五大原则**：禁止运营商封锁网站、禁止减慢加载速度、禁止为加速额外付费、必须增强服务数据透明度和监管无宽带的无线网络。
- FCC颁布的《1934年通讯法案》“公共承运人”为所有用户提供无差别的服务，这一原则延伸到了互联网服务。
- 2017年，FCC宣布网络中立性规则废除。这意味着互联网运营商可以随意控制和减慢一些网站的加载速度，对于依靠视频清晰度和加载速度的流媒体可能造成致命的打击。

网络平台的“避风港原则”

- ➡ **网络服务提供者的“避风港”原则**：提供网络服务的网络公司，对于涉及的内容侵害他人合法权益的，在收到权利人的通知之后，删除涉及产品的，不承担侵权责任。这个原则为网络公司提供了一个免责的安全港湾，因此称为避风港原则。
- ➡ 网络服务提供者只有在知道侵权行为或侵权内容的存在后才有义务采取措施，如删除、屏蔽或是断开链接等。（《民法典》第一千一百九十五条）
- ➡ 明确规定了平台处理“侵权行为”的流程，赋予平台“通知—删除—反通知—恢复”的义务。

抖音与B站：网络暴力一键取证与举报机制

- 面向受害者，抖音提供一键防网暴、互动权限管理等隐私设置能力帮助其阻断网暴路径，并提供一键取证能力，帮助其进行网暴证据留存；
- 针对所有用户，抖音还提供反馈渠道，方便反馈网络暴力信息和行为。
- B站上线防网络暴力专项，发布包含一键取证、防陌生人私信骚扰弹幕优选、消息偏好设置、拉黑举报、一键防护在内的六项功能。B站将对内容进行删除下线，情节严重、影响恶劣的，B站也将酌情对违规账号的使用权限功能等采取部分或整体封禁、对账号添加提醒标识等。

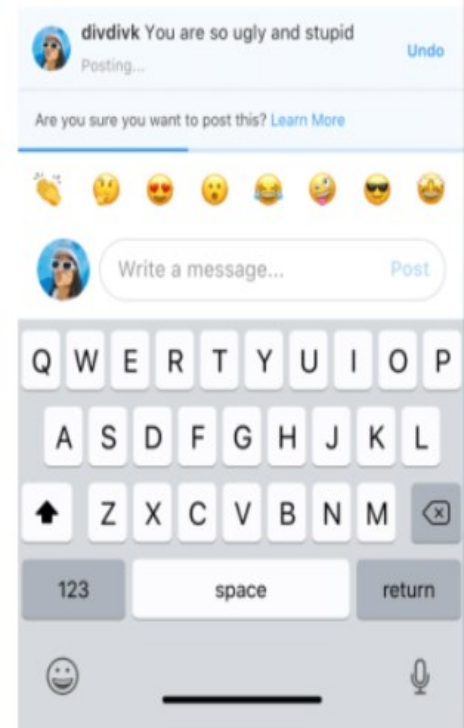


处置效率低：互联网平台的舆情熔断机制

- ➡ 维权周期长
- ➡ 锁定侵权者困难
- ➡ 取证困难
- ➡ 建立恶意话题熔断机制

事前预警：平台的网络暴力检测与过滤

- 面向疑似施暴者，抖音通过站内信、视频推送、弹窗、交互式语音等形式，提醒其注意言论。性质恶劣或屡犯不改的，将依据规范给予相应处罚。对于易出现网暴言论的视频评论区，平台会配置提醒浮条等氛围引导功能，提醒用户理智发声。
- 网络信息服务提供者应当根据历史发布信息、违规处置、举报投诉等情况，**动态管理**涉网络暴力重点账号。



互联网平台的事前预警

- 人工审查、关键词屏蔽
- 网络暴力信息标准，增强识别准确性
- 首次推出了负面评论过滤功能（Instagram）
- 言论提醒和自主选定黑粉功能（Instagram）
- AI技术和人工审核相结合（抖音）
- 评论弹窗警示（抖音）
- 攻击行为冷静期（抖音）

网络暴力识别模型：大数据与算法

- **基于LDA模型和卡方检验的网络暴力话题挖掘方法**：网络语料库中的话题，并用相似度计算方法对话题文本进行分类；然后运用卡方检验筛选话题文本中的暴力特征；最后依据情感词典计算各话题内的暴力值，按照暴力密度判断话题的暴力属性。
- **多模型融合的网络暴力识别**：应用网络爬虫技术爬取受暴人的粉丝列表、微博评论及评论人ID，然后比对施暴评论人ID与受暴人微博粉丝，总结归纳施暴人ID、受暴人微博粉丝与网络暴力的关系（即关注关系），并使用FastText和SVM两个情感分析模型分别对评论内容进行模型训练，最后赋予训练模型和关注关系不同权重，构建网络暴力识别的模型

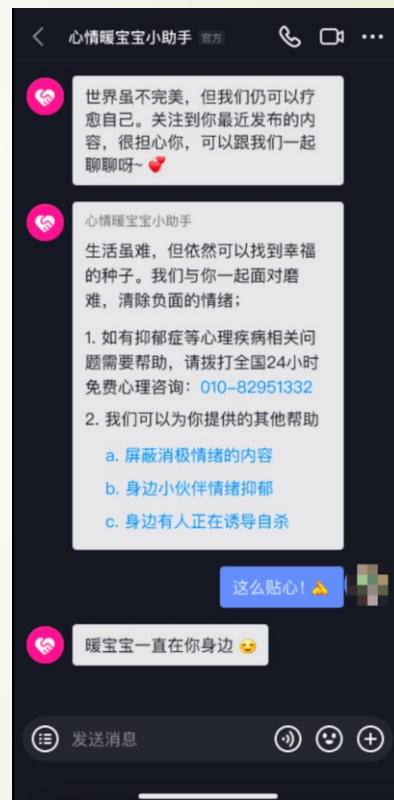
“瓦力”：知乎机器人识别“阴阳怪气”




- 瓦力可通过从情感倾向性、亲密关系、文本特征三方面入手，建立内容情感模型、用户亲密度模型和文本识别模型，并通过训练不断迭代完善
- 实现7个层次的网络反对意见识别，促进和谐友好的社区言论环境构建

心情“暖宝宝”：疏解用户抑郁情绪

如果用户多次违规发布私信、评论，AI算法可以自动匹配触发“心情暖宝宝”，用消息对话尝试缓解用户的抑郁情绪，甚至引导用户去人工求助、线下就诊。





AI识别的技术困境

- ➡ 不擅长理解谁在写或上传图像
- ➡ 文化语境的重要性（意义反差）
- ➡ 谐音、缩写类敏感词有较高的屏蔽率




网络暴力的社会支持

- 物理脱域，避免重复性刺激
- 同伴倾诉
- 父母支持
- 专业心理咨询
- 平台取证
- 网络警察

“一键防护”与“一键取证”：平台责任规定

- 网络跟帖评论、网络社区版块、网络群组、直播和短视频等方面；
- 除了一键防护外，要求平台方提供一键取证，为用户维权，司法机关及有关部门调查取证工作等提供及时必要的技术支持和协助。
- 《规定》第二十条针对三类特殊情形，要求网络信息服务提供者及时协助当事人启动一键防护，切实强化当事人保护。



网络暴力受害者识别与保护

- 用户自述被网暴、被辱骂等场景
- 机器识别
- 医疗机构识别
- 主动发现疑似网暴受害者

“不完美的受害者”和“受害者有罪”理论

- 对受害人非黑即白的“完美”要求，指责有失检点、不反抗等“受害者有罪”论调，是长期以来侵害案件受害人不敢说出真相、放弃指控、造成二次伤害的原因。
- 在“受害者有罪论”的背景下，网络暴力往往被视为受害者自己行为的后果，而非施害者的责任。这种归属理论可以为网络暴力行为提供错误的合理化。

讨论题1

- “雪崩时，没有一片雪花是无辜的”。(No snowflake in an avalanche ever feels responsible)（《伏尔泰语录》）
- 网暴不是一个人能完成的，除了少数发布者之外，大量攻击言论，都来自跟风者，是网民集体非理性行动的结果。
- 网络暴力事件中存在施暴者责任稀释的现象。现实中网络暴力惩治中遵循“责重”原则。


网络暴力“雪崩”发生时，哪些“雪花”应该担责？
如何认识和解决网络暴力事件中的“法不责众”问题？

讨论题2：暴亦有道？

- 在诸多网络暴力事件中，施暴者站在“道德的制高点”对被暴者实施羞辱，形成了多对一的暴力实施情境。
- 从法律意义上讲，网友在进行道德审判时的证据来源、手段、范围都不明确，甚至已经超过了法律的界限。（董晨宇）

网络暴力中的道德审判与法律审判存在何种差异？

规制“道德审判”的可能途径有哪些？



讨论题3：制定《青少年网络暴力防治指南》

- 如何构建青少年网络暴力防治的法律网、技术网和家庭支持网络？
- 试提出你的建议。