Modern Linguistics 现代语言学, 2025, 13(11), 432-441
Published Online November 2025 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2025.13111181

Hans汉斯

网络语言暴力现象及其治理

周小楠，王婷婷

湖北文理学院文学与传媒学院，湖北 襄阳

收稿日期：2025年9月12日；录用日期：2025年11月3日；发布日期：2025年11月17日

摘　要

网络语言暴力现象作为数字时代的新型暴力，严重威胁网络生态与社会文明。本文以网络语言暴力现象中的高频字词为研究语料，聚焦网络语言暴力现象及其治理问题。研究明确了网络语言暴力现象的三重含义，揭示詈语滥用、一般词汇污名化、威胁语与诅咒语、语言异化四类暴力语言的言语特征，明确其对语言规范与社会伦理的双重挑战。最后在遵循语言发展规律的基础上，从政府、社会、个人三个维度提出协同治理方案，以期为构建健康和谐的网络环境提供理论支持和实践指导。

关键词

网络语言暴力，语言特征，语言治理

Cyber Verbal Violence: Phenomena and Governance

Xiaonan Zhou, Tingting Wang

College of Chinese Literature and Media, Hubei University of Arts and Science, Xiangyang Hubei

Received: September 12, 2025; accepted: November 3, 2025; published: November 17, 2025

Abstract

As a new form of violence in the digital age, cyber verbal violence poses a serious threat to the online ecosystem and social civility. This study employs high-frequency words and phrases from instances of cyber verbal violence as linguistic corpora, focusing on the phenomenon itself and its governance. The research clarifies the tripartite definition of cyber verbal violence, revealing the linguistic features of four types of violent expressions: abusive language, stigmatization of common vocabulary, threats and curses, and linguistic alienation. It highlights the dual challenges these phenomena pose to linguistic norms and social ethics. Finally, based on the principles of language development, the study proposes a collaborative governance framework involving government, society, and individuals,

文章引用: 周小楠, 王婷婷. 网络语言暴力现象及其治理[J]. 现代语言学, 2025, 13(11): 432-441.
DOI: 10.12677/ml.2025.13111181

aiming to provide theoretical support and practical guidance for building a healthy and harmonious online environment.

Keywords

Cyber Verbal Violence, Linguistic Features, Language Governance

1. 研究背景

根据中国互联网络信息中心(CNNIC)发布的第 55 次《中国互联网络发展状况统计报告》，截至 2024 年 12 月，我国网民规模达 11.08 亿人，互联网普及率达 78.6% [1]。互联网深度嵌入社会生活的各个维度，网络空间已然成为公众重要的表达场域。然而，网络的开放性特质、虚拟环境以及相对隐秘的交流方式，却也使网络空间成为了某些用户宣泄负面情绪，使用攻击性、暴力性言语的主要场所。“近年来网络暴力事件频发，不断挑战公众认知和法律底线，如何有效防控网络暴力成为极具理论与实践意义的议题”[2]。

其中，网络语言暴力凭借高频次、强伤害性的特点，成为网络暴力的主要形态。这类暴力不仅是现实社会矛盾在网络空间的延伸，更借助文字、符号等载体，将伤害从虚拟世界传导至现实生活。从言语辱骂到恶意造谣，从人身攻击到群体嘲讽，语言暴力的泛滥不仅威胁个体心理健康，还加剧社会信任危机，破坏网络空间的交流秩序。

网络语言暴力现象已突破单纯的交际范畴，演变为影响社会稳定的公共议题，对其进行系统性研究已刻不容缓。因此，本文将立足语言学视角，通过收集网络语言暴力现象中的高频语料，提炼其言语特征进行分析，并在此基础上探讨治理网络语言暴力现象的策略，以期提升社会对网络语言暴力现象的重视程度，维护网络空间的清朗环境。

2. 研究现状及创新

由于网络语言暴力现象涵盖法律、社会学、心理学等多个领域，目前关于网络语言暴力现象的界定，学界尚未达成一致的解释。不同学者之间有不同看法，毛向樱认为：“在自由的网络空间里，利用语言攻击与控制，背离道德标准，对他人的合法权益造成破坏，并且一定程度上侵害他人精神、心理和生理的行为，包含网络谣言、恶意信息的泛滥等多种呈现形式”[3]；刘梦妮则将其定义为：“因某一事件作为主要导火索，以网络媒介为主要载体，对他人进行辱骂、诋毁、诽谤等攻击性的言辞，并且以语言的方式使当事人受到人格和心理的伤害，影响到他人现实生活的行为”[4]。

这表明，当前网络语言暴力研究存在三方面局限：一是概念界定碎片化。国内外学者对网络语言暴力的定义多从单一维度出发(如行为、心理、场景)，尚未形成涵盖“语言形式－行为过程－社会影响”的系统性概念框架，导致研究边界模糊，难以进行跨研究对比。二是技术维度分析缺失。现有研究对语言暴力的特征分析仍停留在传统文本层面，对数字技术催生的新型暴力形态(如 AI 合成图像、编码化詈词)关注不足，无法解释技术赋能下语言暴力的隐蔽性、扩散性新特征。三是治理方案协同性不足。国内研究多侧重法律或平台单一维度，国外研究虽强调技术治理，但缺乏政府、社会、个人的协同设计；且现有方案多为宏观建议，未结合最新政策(如 2024 年《网络暴力信息治理规定》[5])提出可操作的落地路径。

因此，本研究立足语言学视角，在前人研究的基础上实现三方面拓展与创新：

一是创新概念体系。突破单一维度定义局限，首次提出“网络暴力语言－网络语言暴力－网络语言暴力行为”三重含义框架，明确“工具－现象－行为”的逻辑链，填补概念碎片化空白，为后续研究提供统一的概念基准。二是细化网络语言暴力的特征分析。将技术驱动型暴力语言纳入研究范畴，系统分析编码化詈词(如“NMSL”、“404”)、AI 合成图像等新型形态，补充非文本暴力特征的研究，完善中文网络空间语言暴力的特征图谱。三是完善治理框架。结合《网络暴力信息治理规定》[5]最新要求，构建“政府立法执法－社会共识凝聚－个人能力提升”的多主体协同治理体系，既整合了法律、技术、教育等单一治理手段，又通过“智能取证”、“平台预警”、“媒体宣传”的联动设计，解决现有方案协同性不足的问题，为实践提供可操作的路径。

3. 网络语言暴力现象的三重含义

基于此理论背景，本文从语言学视角对网络语言暴力现象进行分层解构，提炼出三重含义：“网络暴力语言”、“网络语言暴力”及“网络语言暴力行为”。三者构成“工具－现象－行为”的逻辑链，既相互依存又各有侧重，共同构成完整概念体系。其一、网络暴力语言：是指在网络环境中使用的，具有攻击性、侮辱性、煽动性、歧视性等特点的言辞。它是一种语言形式，其核心特征是语言符号本身的暴力属性；其二、网络语言暴力：指以网络为媒介，通过文字、图像或语音对特定对象实施的系统性、持续性语言攻击行为，是网络暴力语言在特定情境下的聚集性应用。其本质是“符号暴力”在网络空间的延伸，具有规模性、聚集性和伤害持久性；其三、网络语言暴力行为：指行为主体在网络空间中有意识地使用暴力语言对他人实施攻击的具体动作与过程。它强调的是主体的行为，包括使用暴力语言的意图、方式以及造成的后果等。

三者之间构成“工具－现象－行为”的逻辑链，既相互依存又各有侧重，既有联系又有区别，共同构成了网络语言暴力现象的完整概念。

4. 网络暴力语言的言语特征

网络暴力语言含义丰富，形式多样，涵盖语音、词汇、语义等多个语言要素。我们根据网络暴力语言不同的暴力性内涵，将其分为詈语的滥用、一般词汇的污名化、威胁语和诅咒语、语言异化四类。

4.1. 詈语的滥用

詈语本身就是人类语言中骂人的话，用词粗俗，攻击性强，而互联网的隐蔽性更让网络空间成为詈语俗词的肆虐之地。施暴者通过调用社会文化中的禁忌符号与侮辱性词汇，实现对目标对象的直接攻击。这类词汇因其语义本身的攻击性，在网络空间中普遍传播，主要表现为两类语言模式。

4.1.1. 人格贬损型詈语

以“贱”、“恶”、“坏”、“丑”、“蠢”等粗俗单字或短语为例，其通过极端化语义否定他人人格价值与道德属性。这类词汇凭借简洁且具冲击性的表达，将复杂的社会评价简化为单一的道德审判，进而对受害者的人格尊严、名誉权构成粗暴侵害。

4.1.2. 性语污化

网络暴力语言中有很多涉及性器官、性行为和性关系混乱等语义内涵的词语，统称其为“性语”。其一、与性器官有关的詈语，如“关你卵事”、“你妈个蛋的”、“算个屌”等；其二，与性行为有关的詈语，如“操”、“日”、“肏”等；其三，与性混乱有关的网络詈语，如“他妈的、操他妈、操他奶奶的、操你妈、日你祖宗”[6]。

性语污化本质是将人类生物本能转化为侮辱性符号，打破社会对性话题的禁忌，将生殖器官符号化、污名化，在语言上完成对他人尊严的践踏。借此类词汇试图在心理上对对方形成“我为强者，你为弱者，我能征服你，而你只能顺从我”的压迫与威慑，构建“施害者–受害者”的权力压制关系，满足施暴者的征服欲与心理威慑。

这两类词汇本身具有直接鲜明的攻击性，使得它们即使脱离具体交际场景，仍可以被重复高频使用。当暴力语言突破语境限制，频繁出现在网络空间时，公众对其产生的负面情感会逐渐钝化，进而形成暴力语言“常态化”、“习惯化”的心态，甚至自身也开始使用暴力语言。这种现象不仅消解了语言规范的约束力，更使得攻击性表达在网络空间获得不合理的包容，最终加剧网络语言生态的恶性循环，威胁健康语言环境的构建。

4.2. 一般词汇的污名化

网络语言暴力通过语义的恶意重构，将一般词汇污名化，让许多原本不具有詈骂功能的汉字沦为詈骂语。最典型的是通过原有负面语义放大与新增词汇的负面语义这两种模式，实现负面语义的建立和传播。

4.2.1. 负面语义的极端化放大

在中华传统文化中，“以人为尊，动物为贱”的思想源远流长。因此，在日常生活中，当人们之间产生矛盾冲突时，一方将另一方贬称为牲畜，企图将受害者粗暴地排斥在人类之外，从生物层面对他人进行降格处理，也是一类常见的詈骂表达。

常见的“猪”、“狗”、“蛇”、“龟”等动物词汇，在传统文化概念中，本身带有懒惰、谄媚、阴毒、懦弱等负面属性，而在日常的语言表达中，逐渐衍生出“蠢猪”、“猪脑子”、“哈巴狗”、“狗男女”、“狗娘养的”、“毒蛇”、“王八”、“龟儿子”、“龟孙子”、“龟孙”、“乌龟王八蛋”等詈词。此外，现代语境中新增的“垃圾”、“废物”、“臭虫”这类原本形容低价值或污秽事物的词汇，也被同样地用作贬称詈词。

在网络暴力语境下，这些词汇的负面语义被刻意且极端地放大，把人类与低等生物或无价值事物强行关联，实现对他人整体评价的降格贬损。

4.2.2. 新增词汇的负面语义

网络施暴者常运用语义扭曲与符号重构的手段，将原本中性或褒义的词汇赋予贬义色彩，使之异化为道德谴责的工具。如“白莲花”从清新的植物意象演变成指代“外表清纯，内心阴暗，喜欢装纯洁、装清高的人”的贬称、“绿茶”从日常饮品名称转化为形容“表面清纯无害，实际上城府很深、心机很重的女人”的负面标签；“圣母”原指宗教中象征纯洁崇高的形象(如“圣母玛利亚”)，在网络中被曲解为“不分是非的滥好人”，用于攻击对弱势群体持同情态度者[7]。

这类词汇的负面语义与被攻击对象之间往往无逻辑关联，更没有事实依据，均为施暴者的主观臆断。其暴力实质是借社会刻板印象实施语言暴力，通过散布不实信息损毁他人声誉，达成语言攻击的目的。

4.3. 威胁语和诅咒语

语言是“以言行事”的社会实践，网络暴力语言通过威胁与诅咒两类典型言语行为，以制造恐惧的方式，实现对他人的心理压制与文化规训。

4.3.1. 威胁语

威胁语作为一种言语行为，是指说话者通过语言向特定对象传递不利后果或伤害意图，以此形成心

理压力与威慑，迫使受害者服从施暴者意志或达成特定目的。它的核心意图是制造恐惧和压力，迫使对方屈服。例如，“以后再在网络上看见你，见一次骂你一次”、“你要是再敢发表意见，小心我曝光你的个人信息”、“你作为一个老师，居然敢这么说话，我要把你的话截图发到你们学校，让大家看看你的真面目，让你丢了这份工作”。尽管多数威胁仅停留于言语层面，未引发严重现实后果，但其对当事人造成的强烈心理冲击及恐慌情绪，可能诱发失眠、焦虑等躯体化反应，造成一定程度的现实危害。

4.3.2. 诅咒语

《现代汉语词典》(第 6 版)关于“诅咒”的释义是：“原指祈祷鬼神加祸于所恨的人，今指咒骂”[6]。在网络暴力语言中则泛指各种形式的恶意咒骂。它们通过话语传递伤害性意图，利用文化禁忌强化伤害。其内容与形式多样，主要可归纳为以下三类。

其一为生命诅咒类，如“绝症缠身”、“不得善终”等表述，人的生命只有一次，并且高于一切，因此大多数人都非常注重自己身体状况，以求健康长寿。这类生命诅咒语直接冲击人们的死亡禁忌，借人类对疾病与死亡的本能恐惧制造强烈的心理压迫；其二为家族贬损类，以“断子绝孙”、“家族凋敝”等涉及血缘传承的诅咒为典型，其危害范围从个体延伸至家族。常言道“不孝有三，无后为大”，这类诅咒正是触动了中国传统的家族观念与传承意识，使受害者承受文化传统层面的精神重压；其三为社会否定类，针对事业、学业等社会价值，如“考试失利”、“早晚破产”等预言，通过否定对方的社会角色与劳动价值，虚构失败前景以制造心理阴影。

从诅咒语的伤害途径分析，网络诅咒语构建了双重暴力：一方面通过负面语言直接施加心理压力，满足施害者的报复性宣泄；另一方面巧妙运用传统文化的禁忌避讳，使受害者从侧面遭受习俗文化上的隐性压迫。这种语用机制不仅加剧了语言伤害的强度，更凸显了网络语言暴力对社会文明与伦理秩序的深层挑战。

4.4. 语言异化

网络传播的互动性促使网民积极参与热点事件的讨论，同时媒介的匿名性、即时性和低门槛特性，使网民能够隐藏真实身份，轻易参与暴力传播，将“吃瓜”心态转化为“玩梗”式伤害，形成特定的网络“黑话”，产生新的网络暴力语言形态。

4.4.1. 编码化詈词

在网络传播的监管日渐严格的压力下，部分网民为满足个人情绪宣泄的目的，逐渐创造出利用拼音首字母、数字谐音或数字字母混合等新型詈词形态。例如：BT (变态)、“SSFD”(瑟瑟发抖，用于嘲讽胆怯)、“NT”(“脑瘫”)、“NMSL”(原意为“你妈死了”，后异化为攻击性符号)；740 (气死你)、0748 (你去死吧)、“404”(网页错误代码)被借指“消失”，衍生出“祝你人生 404”；2CS (俩畜生)、3QS (仨禽兽)等缩略词、谐音词。此类表达通过编码化处理，模糊了这类语言的暴力本质，使一般人难以分辨，从而构建起暴力传播的“隐形通道”，逃避了平台算法的监管。

4.4.2. 特定语境下的詈骂语

“随着社会的发展及文化的变迁，人们会形成新的表达习惯，许多詈骂语中可能不含有詈词，但仍然含有詈骂的性质，这种情境下的詈骂含义无法直接通过某个单一语素表达出来，它的詈骂性质需要结合语篇、语境传达，整个句子才会构成詈骂属性”[8]。这就是在网络语言暴力语境中以“冷嘲热讽”为表现形式的隐性詈骂语。它们以“合理化质疑”为伪装，通过语境绑定实现精神攻击，其伤害性往往比直接辱骂更持久。

以“武汉校园碾压案”孩子妈妈不堪网暴跳楼离世的案件为例。在该案例中众多网民并没有使用辱

骂、威胁等具有明显人身攻击性的词语，而是刻意脱离其丧子之痛的核心语境，将矛头转向对受害者外在表现的无端指摘。诸如“还能穿这么正式？”、“这是孩子妈妈吗？说话为何如此冷静？”、“妈妈的穿着打扮是用了心的”、“这位妈妈想成为网红吗？”[9]等言论，表面上虽不构成直接辱骂，实则构成一种具有讥讽意味的话语攻击。可在网络暴力的特殊语境之下，这些言论如同一把把软刀子，对这位母亲的心理造成了极大伤害。“而且在此时调侃一个母亲失去儿子的痛苦，在本质上无疑具有强烈的人身攻击性，最终导致这位母亲跳楼自杀的严重后果”[9]。

“因此，判断网络语言是否具有人身攻击性要从形式和实质两方面入手”[9]。对于表面上没有人身攻击性，但在网络暴力特殊语境中具有人身攻击效果的语言，并对受害者造成心理压迫与精神伤害，就应当被认定为网络语言暴力的实施手段之一，并纳入相应的规制范畴。

4.4.3. 表情包与图像暴力

从 P 图软件的兴起到现在 AI 技术的发达，网络媒介将语言暴力从文本维度拓展至视觉符号系统。部分网民通过制造表情包、动图、AI 合成内容等形式对他人实施网络暴力。例如截取影视片段中的丑化表情(如“翻白眼”、“狰狞脸”)，配以“你算什么东西”等文字。这种刻意丑化的表情包，将人物神态异化为攻击武器，在互动中实现对他人的实时羞辱；或者制作带有“绿茶”、“圣母”标签的讽刺表情包，并在聊天、评论中高频传播，强化了这类词语的负面标签；更有甚者通过 AI 技术合成虚假图像，如“换脸”、“语音克隆”，将受害者形象嵌入色情、暴力场景，实施“技术性造黄谣”。2023 年杭州互联网法院审理的“粉色头发女孩被 AI 造黄谣案”中，加害者通过篡改图像传播谣言，导致受害者陷入“社会性死亡”境地，最终自杀身亡。

此类现象的危害性体现在：一方面，表情包把暴力包装成“玩笑”、“娱乐”，让人容易忽视其中的伤害性；另一方面，AI 合成的虚假图像能逼真地伪造“事实”，比文字更有迷惑性，伤害也更大。视觉形式的暴力不仅模糊了“开玩笑”和“攻击”的界限，更因为图像传播更快、影响更广，极大地增加了受害者澄清事实与依法维权的难度。

总体而言，网络暴力语言呈现詈语的滥用、一般词汇的污名化、威胁语和诅咒语、语言异化的多维乱象：一方面，暴力词汇的滥用、中性词语的恶意污名化以及威胁诅咒等不当的汉字用语，直接违背汉语的词汇语义规范和交际准则，破坏语言的纯洁性与规范性；另一方面，媒介技术的发展催生出中英文混用、字母缩写滥用、表情包与图像暴力等新型形态，让暴力语言更具隐蔽性与迷惑性，模糊了正常表达与语言暴力的边界。这些乱象不仅对个体造成精神伤害，更污染着网络语言生态。因此，强化网络暴力语言治理，营造清朗健康的网络语言环境迫在眉睫。

5. 网络语言暴力的解决措施

自 2024 年 8 月 1 日起，国家互联网信息办公室联合公安部、文化和旅游部、国家广播电视总局四部门出台的《网络暴力信息治理规定》[5] (以下简称《规定》)正式施行。该《规定》从明确网络信息内容管理主体责任、建立健全防范预警机制、规范网络暴力信息处置流程、加强用户权益保护、强化监督管理、明确法律责任等多个维度，为网络暴力治理提供了基础性制度框架。本文以《规定》为理论与政策依据，从政府、社会、个人三个层面展开分析，重点围绕“平台责任”这一核心环节深入探讨治理难点与细化路径，同时兼顾其他层面的对策优化，旨在构建更具可操作性的多维度网络语言暴力治理体系。

5.1. 政府层面：立法与执法双管齐下

5.1.1. 专项立法的精细化完善

当前我国网络语言暴力治理面临“泛法可依但专法缺失”的困境。尽管《网络安全法》《个人信息保

护法》《治安管理处罚法》《民法典》《刑法》等法律法规，能为受害者提供一定的维权依据，但针对网络语言暴力的专项法律仍为空白[10]，使得网络语言暴力的行为界定、责任划分、量刑标准等缺乏专项条款，导致司法实践中存在三大问题：一是“暴力言论”与“正常批评”的边界模糊，如对公共事件的争议性评价是否构成“侮辱”、“诽谤”，不同法院的认定标准存在差异；二是“情节严重”的认定缺乏量化指标，现行法律未明确网络暴力的传播次数、影响范围、造成的精神损害程度等具体标准，导致部分案件因“情节认定不清晰”难以立案；三是责任主体认定复杂，当暴力言论经多平台转发、匿名账号传播时，责任追溯往往止步于“平台删除”，难以追究源头施暴者责任。

基于此，专项立法需从三方面细化：其一，明确网络语言暴力的“行为清单”，结合《规定》中“网络暴力信息”的定义，进一步列举“人肉搜索”、“恶意 P 图”、“持续性辱骂”、“煽动群体对立”等具体行为，并区分“轻微违规”、“一般违法”、“刑事犯罪”三个层级的认定标准；其二，设定“情节严重”的量化指标，如规定“同一暴力言论在单一平台传播超 500 次”、“造成受害者出现心理疾病诊断证明”、“引发线下骚扰或人身威胁”等情形，直接作为立案或从重处罚的依据；其三，建立“责任追溯链条”，明确平台需留存施暴账号的注册信息、IP 地址、登录设备等数据至少 6 个月，若因平台数据留存不全导致责任无法追溯，需承担补充赔偿责任。

5.1.2. 智能取证体系的实操性构建

目前，受害者在遭遇网络语言暴力后，通常只能通过报案或自诉来为自己维权，但这两种方式均需要通过充足的证据来证明犯罪行为。与传统违法犯罪不同，网络语言暴力恰恰是以精神伤害为主，是非接触性的，而精神损害缺乏直观鉴定依据，现行伤情鉴定体系多针对身体伤害，受害者需提供心理咨询记录、医院诊断证明等间接证据，证明难度大，这就导致受害者的伤情难以鉴定；同时，网络语言暴力往往针对素不相识的陌生人实施，以电子证据为主，不同于传统的实质证据，电子证据不仅取证困难，且真假难辨[10]。电子证据易篡改、易灭失，施暴者可通过注销账号、使用虚拟 VPN、修改数据等方式逃避追诉，公安机关传统取证手段难以应对海量网络数据筛选需求。因此，网络语言暴力的“非接触性”与“电子证据特性”，导致受害者维权面临“取证难、鉴定难、溯源难”三大现实障碍。受害人和公安机关往往在伤情鉴定、确认侵害人、收集证据等方面存在现实困难，维权成本极高。

针对上述问题，公安机关需构建“技术 + 协作”双驱动的智能取证体系：在技术层面，开发“网络暴力电子证据固定系统”，整合大数据筛选、区块链存证、AI 溯源三大功能——通过大数据技术对同一事件下的暴力言论进行聚类分析，快速定位传播热度高、影响范围广的核心内容；利用区块链技术对电子证据(如截图、聊天记录、转发记录)进行固定，确保证据不可篡改；通过 AI 算法关联分析施暴账号的登录设备、常用 IP、语言习惯等特征，识别匿名账号背后的真实用户。在协作层面，建立“公安－网信－平台”实时数据对接机制，根据《规定》要求，平台需在接到公安机关取证请求后 24 小时内，提供涉案账号的注册信息、传播路径等数据，同时公安机关需定期向平台反馈取证需求类型，帮助平台优化数据存储与调取流程。此外，联合高校开设“网络安全与电子取证”专项课程，培养兼具法律知识与技术能力的复合型取证人才，解决基层公安机关技术人才短缺问题[10]。“同时强化衔接配合。人民法院、人民检察院公安机关要加强沟通协调，统一执法司法理念，有序衔接自诉程序与公诉程序，确保案件顺利侦查、起诉、审判”[11]。

5.2. 社会层面：凝聚共识和平台治理

5.2.1. 媒体：从“单向宣传”到“双向共识构建”

媒体作为信息传播与舆论引导的关键角色，在网络语言暴力治理中肩负重任。当前网络语言暴力治理尚未形成全民共识，部分网民将“攻击性言论”等同于“言论自由”，对网络暴力的现实危害认知不

足。媒体需打破传统单向传播模式，通过“案例解读 + 公众参与”的方式构建治理共识，构建抵制网络语言暴力的社会心理基础，从而塑造理性网络生态。

一方面，针对热点问题或公共性的热点事件，新闻媒体可以约请专家开展专题研究，引导形成全民关注、全民参与的舆论场[12]。例如聚焦“杭州粉发女孩被网暴”、“寻亲少年刘学州”、“杭州女子取快递被造谣出轨案”等典型案例，不仅报道事件经过，更需联合心理学专家、法律从业者解读“网络暴力如何从虚拟言论转化为现实伤害”，通过报道揭示暴力言论从网络空间向现实生活的传导路径，引导公众认识到网络语言暴力并非虚拟世界的“无害游戏”，而是可能引发严重现实后果的侵权行为。另一方面，开设“网络文明议事厅”专栏，构建与公众的对话通道，通过短视频平台、社区论坛等渠道，邀请网民分享自身遭遇或目睹的网络暴力经历，提出治理建议，同时将公众建议整理成《网络暴力治理公众意见报告》，提交至网信部门与平台，形成“宣传–反馈–优化”的闭环。此外，媒体还可加强权威信息披露，通过曝光网络语言暴力的典型案例与惩处结果，发挥舆论监督作用，形成对潜在施暴者的震慑效应，引导网民理性发声，推动全社会形成抵制网络语言暴力的价值共识。

5.2.2. 平台建立网络语言暴力预警系统

2022 年中央网信办“清朗 • 网络暴力专项治理行动”[13]，要求平台构建“预警–干预–保护”全链条机制，但实践中仍存在 AI 识别准确率不足、分级响应僵化、用户保护功能低效等问题，具体优化方案如下：

一是优化 AI 识别技术：平衡“精准性”与“包容性”。当前平台 AI 识别系统存在两大核心问题：一是“误判率高”，将正常争议言论、小众群体表达(如亚文化圈层用语、地方方言)误判为暴力内容，如某平台曾将用户“这个政策落地有难度，还需优化”的评论标记为“负面攻击”；二是“漏判率高”，对隐晦性暴力言论(如谐音梗辱骂、隐喻式人身攻击)识别能力不足，如“你这种人就该‘消失’”这类未直接使用脏话的言论，往往无法被系统捕捉。

针对上述问题，平台需从三方面优化技术：其一，构建“多维度识别模型”，突破传统“关键词匹配”的单一模式，增加“语境分析”、“情感倾向”、“用户关系”三个维度——通过语境分析判断言论是否处于争议讨论场景，如在“公共政策讨论”中，对“批评性言论”的判定标准适当放宽；通过情感倾向分析区分“理性反驳”与“恶意攻击”，如同样是“反对观点”，带有“愤怒”、“厌恶”等强烈负面情绪的言论需重点审核；通过用户关系分析判断是否存在“针对性霸凌”，如陌生账号对同一用户的持续性评论，需提高识别敏感度。其二，建立“小众言论样本库”，收集亚文化用语、方言、专业领域术语等非通用表达，标注其正常含义与暴力语境下的差异，如“卷”在职场语境中为正常表述，但若用于“你这么卷，难怪没人喜欢你”则带有攻击意味，通过样本训练提升 AI 对小众言论的识别精度。其三，设置“人工复核绿色通道”，对 AI 判定为“疑似暴力内容”的言论，若用户提出申诉，需在 12 小时内由专业审核人员(兼具语言分析与法律知识)进行复核，同时定期统计 AI 误判案例，反哺模型优化，将误判率控制在 5%以下。

二是建立分级响应机制，避免“一刀切”。部分平台现行分级响应机制存在“标准僵化”问题，如无论暴力言论的传播范围、影响程度如何，均采取“删除内容 + 警告账号”的统一措施，导致治理效果大打折扣。结合《规定》要求，平台需建立“风险等级–处置措施”动态匹配机制，具体而言，可将网络暴力言论划分为低、中、高三个风险等级，并对应不同的判定标准与处置措施。其中，低风险言论指无明显辱骂、诽谤内容，仅含轻微不文明用语，且传播范围限于个人主页或小范围好友圈、未引发互动的单一言论，平台应对其标注“不文明用语提醒”，并向用户发送“文明发言指引”私信；中风险言论指含明显辱骂、讽刺内容(未涉及人身攻击或隐私泄露)，或传播至平台热门话题区且转发或评论数达 50~500 次，

或引发少量负面互动的言论，平台需删除违规内容，对账号限制评论功能 7 天，并记录违规次数，累计 3 次则升级处置；高风险言论指含人身攻击、诽谤、人肉搜索、煽动对立等内容，或传播至平台首页、跨平台且转发或评论数超 500 次，或已造成受害者投诉、心理伤害、线下骚扰的言论，平台需立即删除内容并屏蔽相似信息，永久封禁涉事账号且上报网信部门备案，同时向受害者提供“一键取证”工具，留存证据供公安机关调取。同时，平台还需建立“风险等级动态调整机制”，若低风险言论在 24 小时内传播范围扩大、互动量激增，需自动升级为中风险；若高风险言论经处置后仍有其他账号转发，平台需启动“相似内容全网排查”，避免暴力言论“换号传播”。

三是升级用户保护功能，从“有功能”到“好用”。当前部分平台的用户保护功能存在“操作复杂”、“响应缓慢”问题，如“一键防护”功能需用户手动开启多个子选项(限制陌生人评论、隐藏个人信息等)，“快速举报通道”提交后需 3~5 个工作日才能收到反馈，导致用户使用率低。平台需从“用户体验”出发优化功能：其一，简化“一键防护”操作，用户开启后自动激活“陌生人评论过滤”、“个人信息(手机号、住址)隐藏”、“暴力言论自动屏蔽”三项核心功能，同时提供“自定义防护”选项，满足不同用户需求；其二，升级“快速举报通道”，对“网络暴力”类举报设置专属审核队列，承诺 2 小时内反馈初步处理结果(如“已受理，正在核查”)，24 小时内反馈最终处置结果(如“违规内容已删除，账号已警告”)；其三，新增“受害者心理支持”功能，在用户提交网络暴力举报时，同步提供专业心理咨询机构的联系方式与免费咨询名额，帮助受害者缓解心理压力。

四是加强热点事件舆情管控：避免“次生暴力”。热点事件往往是网络语言暴力的高发场景，部分平台因“信息审核滞后”、“权威信息展示不足”，导致暴力言论快速扩散。平台需建立“热点事件专项管控机制”：在事件发酵初期(热度上升至平台热搜前 50 名时)，启动“信息传播时间差”策略，对非权威来源(如个人账号、非认证媒体)的信息设置 15 分钟延迟发布，同时优先展示政府部门、主流媒体发布的权威信息，避免谣言引发对立；在事件发酵中期，组建“热点事件审核专班”，实时监控评论区、话题区的暴力言论，对批量出现的类似暴力内容，启动“批量屏蔽 + 账号溯源”；在事件平息后，发布《热点事件网络暴力治理报告》，公开处置的违规账号数量、典型案例，引导网民理性看待热点事件。

5.3. 个人层面：提升语言能力和道德修养

网络空间作为当代语言实践的新兴场域，个体作为语言使用的最小单元，其语言选择与语用行为构成网络语言环境的微观基础。

首先，个体应强化语言规范意识，自觉学习语言文字规范标准，遵循语言文字的使用和书写规则。通过对比正常交际话语与暴力话语的语法、语义及语用差异，明晰语言暴力的边界。部分网民因缺乏对“网络语言暴力边界”的认知，无意识地发表攻击性言论。需通过“对比教育 + 场景模拟”的方式提升认知：一方面，学校、社区可制作《网络语言规范手册》，对比“正常交际话语”与“暴力话语”的差异，如“我不认同你的观点，理由是……”属于理性表达，“你这观点太蠢了，根本没脑子”则属于暴力言论，通过具体案例帮助网民明晰边界；另一方面，在短视频平台、社交 APP 中嵌入“语言模拟测试”小游戏，设置“公共讨论”、“私人聊天”等场景，让用户判断不同言论是否构成网络暴力，测试完成后提供解析与规范建议，增强学习趣味性。“同时要提高对信息的筛选能力、质疑能力、理解能力、评估能力，做到理性发声；还应学习愤怒情绪管理艺术，不要肆意宣泄、过度发泄”[14]。

其次，单纯的“法律条文宣传”难以让网民形成深刻认知，需结合具体场景普及法律后果：学校可开设“网络法律实践课”，通过模拟“网络暴力案件庭审”，让学生分别扮演法官、原告、被告，直观了解“发表暴力言论可能面临的民事赔偿、行政处罚甚至刑事责任”；社区可邀请经历过网络暴力维权的受害者分享案例，讲解“如何收集证据、向哪个部门投诉、维权流程有哪些”；网络媒体可制作“网络暴

力法律后果”短视频，使网民了解网络语言暴力的法律后果，增强自我约束意识。

此外，鼓励公众参与网络语言生态的共建共治，支持网民通过平台举报、舆论监督等方式，对网络语言暴力行为进行主动干预，如平台可设立“网络文明志愿者”机制，对积极举报网络暴力、传播文明言论的用户，给予“积分奖励”(可兑换平台会员、公益捐赠资格等)，同时定期公示“优秀志愿者名单”，激发参与热情，推动网民从网络语言暴力的“被动受众”转变为网络文明生态的“建设主体”。

语言系统本身具备动态平衡的调节机制，但网络暴力语言作为数字时代的语言生态失衡现象，已然超出自然净化的阈值。我们需要在遵循语言发展规律的前提下，对网络暴力语言现象进行有效的规划和治理，兼顾技术理性与人文关怀，形成政府、社会和个人构建的多主体协同治理体系。让语言回归理性表达本质，实现网络语言生态的动态平衡。

基金项目

湖北文理学院大学生创新创业训练项目“网络暴力语言现象及其治理”(S202410519070)成果。

参考文献

[1] 第 55 次《中国互联网络发展状况统计报告》发布[J]. 传媒论坛, 2025, 8(2): 121.

[2] 全国信息网络安全协会联盟. 2022 年全国网民网络安全感满意度调查专题报告：网络暴力防控与网络文明专题[EB/OL]. 2023-01-04. https://iscn.org.cn/uploadfile/2023/0104/9.pdf, 2025-05-18.

[3] 毛向樱. 网络语言暴力行为的社会交往分析[J]. 哈尔滨师范大学社会科学学报, 2018, 9(1): 108-111.

[4] 刘梦妮. 网络公共事件中的语言暴力现象研究[D]: [硕士学位论文]. 武汉: 湖北大学, 2019.

[5] 国家互联网信息办公室, 中华人民共和国公安部, 中华人民共和国文化和旅游部, 国家广播电视总局. 网络暴力信息治理规定[EB/OL]. https://www.cac.gov.cn/2024-06/14/c_1720043894161555.htm, 2024-06-12.

[6] 李舒慧. 网络暴力语言现象探析[D]: [硕士学位论文]. 锦州: 渤海大学, 2013.

[7] 任仝. 网络语言暴力现象分析[D]: [硕士学位论文]. 长春: 吉林大学, 2019.

[8] 李敏婕. 网络詈词分类分级及其监管应用研究[D]: [硕士学位论文]. 武汉: 武汉大学, 2019.

[9] 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.

[10] 孙钰琨. 网络语言暴力的现状、成因及解决措施研究[C]//百色学院马克思主义学院, 河南省德风文化艺术中心. 2023 年高等教育科研论坛南宁分论坛论文集. 北京: 中国人民公安大学, 2023: 268-270.

[11] 最高人民法院, 最高人民检察院, 公安部. 关于依法惩治网络暴力违法犯罪的指导意见[EB/OL]. http://gongbao.court.gov.cn/Details/12dfb372281fcfc26a1489d012108b.html, 2023-09-20.

[12] 徐默凡. 网络语言暴力的界定方法和治理对策[J]. 青年记者, 2023(17): 75-76.

[13] 中央网信办. 中央网信办部署开展“清朗·网络暴力专项治理行动” [EB/OL]. 2022-04-24. https://www.cac.gov.cn/2022-04/24/c_1652422681278782.htm, 2025-05-18.

[14] 王晓燕. 网络语言暴力中的修辞现象探析[J]. 修辞研究, 2022(2): 162-173.