

# 网络暴力侵权中平台过错的解释论重构

贺 平

天津大学法学院，天津

收稿日期：2025年12月7日；录用日期：2026年1月13日；发布日期：2026年1月23日

---

## 摘要

在网络暴力侵权中，平台过错的认定面临解释论困境。传统的“通知 - 删除规则”因被动受理模式难以应对群体性、大规模的网络暴力，导致受害人救济不足。有观点主张引入《民法典》第1198条的安全保障义务以实现网络暴力的事前预防，但该路径存在空间属性解释障碍及补充责任救济效果有限等瑕疵。2024年国家网信办发布的《网络暴力信息治理规定》为构建平台的普遍审查义务提供了技术规范基础。本文主张，应以普遍审查义务对《民法典》第1197条“知道规则”中的注意义务标准进行解释论重构，通过分阶段设定审查范围与标准，实现事前预防与综合治理目标，并借助公益诉讼等公私法协同机制，有效落实平台在网络暴力治理中的责任。

---

## 关键词

网络暴力侵权，平台过错，普遍审查义务

---

# Reconstruction of the Legal Dogmatic of Platform Fault in Online Violence Infringement

Ping He

School of Law, Tianjin University, Tianjin

Received: December 7, 2025; accepted: January 13, 2026; published: January 23, 2026

---

## Abstract

In cases of online violence infringement, the determination of platform fault faces interpretative dilemmas. The traditional “notice-and-takedown rule” struggles to address group-based and large-scale online violence due to its passive processing model, leading to insufficient relief for victims. Some viewpoints advocate for introducing the safety and security obligations under Article 1198 of

**China's Civil Code to achieve preemptive prevention of online violence. However, this approach has flaws, including interpretive challenges regarding spatial attributes and the limited remedial effectiveness of supplementary liability. The "Provisions on the Governance of Online Violence Information", issued in 2024 by the Cyberspace Administration of China, provides a technical regulatory basis for establishing a universal review obligation for platforms. This paper argues for an interpretative reconstruction of the duty of care standard within the "knowledge rule" of Article 1197 of the Civil Code, using a "universal review obligation" framework. This reconstruction would involve setting phased review scopes and standards to achieve the goals of preemptive prevention and comprehensive governance. Furthermore, leveraging public-private law coordination mechanisms, such as public interest litigation, could effectively implement platform responsibility in the governance of online violence.**

## Keywords

**Online Violence Infringement, Platform Responsibility, Universal Review Obligation**

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

### 1. 问题的提出

根据中国互联网络信息中心(CNNIC)于 2023 年发布的第 52 次《中国互联网络发展状况统计报告》显示，截至 2023 年 6 月，我国网民规模达 10.79 亿人。互联网普及率高达 76.4%。互联网的深度普及虽然极大地推动了社交媒体应用的繁荣，但也为网络治理带来了严峻难题，尤其是屡禁不止且大有愈演愈烈之势的网络暴力问题，已经成为了社会的重要关切。

根据中央网信办于 2022 年发布的《关于切实加强网络暴力治理的通知》，网络暴力是指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，侵害他人合法权益，扰乱正常网络秩序的行为。相较于一般的网络侵权行为，网络暴力侵权具有高度的扩散性，能够在短时间内发展成为参与人数众多的群体性侵权，一旦不能及时制止就容易造成无法弥补的损害后果。平台作为连接施暴者和受害者的关键媒介，责任不容忽视。中央网信办于 2021 年发布的《关于进一步压实网站平台信息内容管理主体责任的意见》中指出，维护良好网络生态，需要进一步压实网站平台信息内容管理主体责任，充分发挥好网站平台信息内容管理第一责任人作用。

在既有的网络服务提供者的侵权规则体系下，平台的过错难以证成。通说认为，“通知 - 删除规则”是网络服务提供者过错认定的原则性规范，平台只要在接到侵权通知后采取措施即可免责。然而从网络暴力治理的角度来看，传统的被动受理模式难以应对网络暴力攻击规模海量性、指数级传播快速性和跨平台传播流动性的特征，“通知 - 删除规则”的义务标准失之过宽。有观点认为，可以通过《民法典》第 1198 条的安全保障义务条款对平台的注意义务进行改造，在平台未对受害人采取合理保护措施时承担侵权责任，其虽然在一定程度上提供了注意义务标准，但也面临着从“物理场所”到“网络场所”的解释论困境，和补充责任条款所带来的救济路径上的重大瑕疵。

相较而言，社交平台主动采取措施的“知道规则”更有利于保护受害人，“知道或者应当知道网络用户利用其网络服务侵害他人民事权益”的规范表达也为普遍审查义务的引入提供了规范空间。最高人民法院于 2014 年发布的《关于审理利用信息网络侵害人身权益民事纠纷案件适用法律若干问题的规定》中，将“网络服务提供者采取预防侵权措施的技术可能性及其是否采取了相应的合理措施”列为认定网

络服务提供者是否“知道”的因素之一。我国正在加快网络暴力治理的相关立法，通过网络暴力治理技术标准的引入，国家网信办于2024年6月公布的《网络暴力信息治理规定》中对平台提出了一系列“反网暴”技术要求，为普遍审查义务引入“知道规则”提供了方向与技术上的指引。

本文将基于解释论立场，对“通知-删除规则”与安全保障义务条款的局限与不足进行说明，并对普遍审查义务引入“知道规则”的理论基础与技术实现路径进行论证，以期为平台的过错认定提供更为现实有效的解释论方案。

## 2. “通知-删除规则”的解释论困境

### 2.1. 作为核心规范进路的“通知-删除规则”

“通知-删除规则”和“知道规则”作为我国网络平台的间接侵权责任专款，于2006年颁布的《信息网络传播权保护条例》被首次提出，其主要目的是为了应对网络版权侵权领域中平台不作为的问题。我国于2009年出台的《侵权责任法》将“通知-删除规则”和“知道规则”的适用范围从版权侵权领域拓展到了网络侵权的其他领域，并在《民法典》中得到了延续。

所谓“通知-删除规则”即依据《民法典》第1195条的规定，当网络用户利用网络服务实施侵权行为时，被侵权人有权通知平台采取删除、屏蔽、断开链接等必要措施，若未能及时采取必要措施的，网络服务提供者对损害的扩大部分与该网络用户承担连带责任。只要权利人发出合格通知，就触发了平台的被动履行义务，网络平台接到通知后应当及时根据侵权的初步证据和服务类型采取必要措施，以尽可能避免相关信息进一步传播。就属性而言，该规则首先是网络服务提供者的免责抗辩，即原则上只有当受害人向网络服务提供者发出通知，网络服务提供者才应采取必要措施，网络服务提供者本身并无一般性的事先审查义务；其次，该规则还意味着一旦受害人发出通知，而网络服务提供者未采取必要措施，不论其不作为是出于故意不理睬、过失未注意，还是经过评估认为不成立侵权，并且最后结果证明侵权的确存在，那么，网络服务提供者一律需要承担间接侵权责任。

我国学界的主流观点认为“通知-删除规则”是网络侵权中平台间接侵权责任的核心规范进路，而“知道规则”则属于“通知-删除规则”的例外情形。所谓“知道规则”即《民法典》第1197规定的，网络服务提供者知道或者应当知道网络用户利用其网络服务侵害他人民事权益，未采取必要措施的，与该网络用户承担连带责任。典型观点如张新宝教授所指出的，将“知道规则”例外化的必要性在于该条款会架空“通知-删除规则”的适用，并最终导致平台责任的过度扩张<sup>[1]</sup>。“知道”包括“明知”和“应知”两种主观状态，在以“通知-删除规则”为一般适用条款的情形下，只有少数明显程度极高和社会影响程度极大的侵权事件才能作为“明知”和“应知”的辐射范围。若将解释论上的主次序位颠倒，“明知”和“应知”的涵义外延将得到极大的扩张，被侵权人多会直接诉诸“知道规则”，“通知-删除规则”则会形同虚设，显然有违立法者的本意。

“通知-删除规则”在网络暴力侵权中面临着“水土不服”的问题，其预想功能在既有的网络暴力治理实践中频频“失灵”，在司法实践中也鲜有适用，以“通知-删除规则”为核心规范进路的解释论体系备受质疑。

### 2.2. 规范溯源：《数字千年版权法》中的“避风港原则”

“通知-删除规则”滥觞于美国《数字千年版权法》的“避风港原则”，根据《数字千年版权法》第512条“与在线材料有关的责任限制”规定，网络服务提供者在以下条件下不承担经济赔偿责任：一是网络服务提供者没有实际知道侵权行为的情况；二是在知道该情况后，网络服务提供者迅速采取删除或禁止访问等必要措施；三是信息储存服务提供者和信息定位服务提供者没有从侵权活动中获得营利。

在《数字千年版权法》出台之前，美国的司法实践对于涉及平台的版权侵犯案件通常是以直接侵权规则认定平台企业的版权责任。典型案例例如 1993 年的“Playboy Enterprises v. Frena”案，一些网络用户未经 Playboy 公司许可，在 Frena 论坛上使用了该公司出版杂志中的上百幅图片。法院认为，不仅这些网络用户构成直接侵权，作为网络服务提供者的 Frena 论坛也构成直接侵权，因为“已有证据表明存在着直接的著作权侵权，即使被告有可能不知道著作权侵权的情况，也并不影响侵权的成立”。但是“直接侵权责任说”的观点在后来的司法实践中受到了否定，如 1995 年的“Religious Technology Center v. Netcom On-Line Communication Services, Inc.”案中，法院认为，Playboy 案的结论并不合理，网络服务提供者的工作对于互联网的正常运行而言是必不可少的，并且在整个网络的服务器上，因此几乎不可能从“非侵权比特”选出“侵权比特”，因此采用可能导致无数当事方承担责任的规则是没有意义的，作为网络服务提供者的被告不应承担直接侵权责任。但是法院也表示，如果被告在收到原告的通知后未采取及时措施，或从用户的侵权活动中获得了直接的经济利益，则需要承担间接侵权责任。在 Religious Technology Center 案之后的案件接受了这一主张，并逐渐演变为《数字千年版权法》的“避风港原则”。

“避风港原则”既是“间接侵权责任说”的理论转向的产物，亦是美国网信企业和版权利益集团博弈的结果，在大量平台企业因面临巨大潜在版权责任而求助于国会的同时，美国主要版权行业和软件生产公司也在努力改变平台企业的版权责任，要求平台企业主动控制未经授权的内容流动，以防止美国的平台经济发展给版权行业带来毁灭性的打击[2]。其立基于二十一世纪初网信产业和版权行业的利益分配格局，较好地平衡了版权保护与网信产业发展之间的关系。在“避风港原则”的支持下，网络服务提供者往往以被动性、工具性和中立性自居，声称仅提供“通道和技术支持”，未对网络用户发布的信息进行编辑、修改、筛选，不应承担普遍的审核义务[3]，该观点又被概括为网络服务提供者的“技术中立”。

然而值得注意的是，“避风港规则”源自对 20 世纪末美国版权侵权的判例总结，其规范张力能否克服时代以及适用范围上的局限性，网络服务提供者的“技术中立”态度是否能满足信息爆炸时代的网络治理要求，显然不无疑问。美国版权局于 2020 年发布的《数字千年版权法》评估报告中指出，“今天的互联网与 1998 年颁布《数字千年版权法》时的互联网截然不同”，“随着在线活动和第三方对创作者内容的使用增加，‘避风港原则’满足所有利益相关者需求的能力也面临压力”[4]。网络暴力侵权是信息爆炸时代的产物，在《数字千年版权法》颁行的时代，线上社交的用户规模有限，且多集中于 BBS 电子公告系统、即时聊天室等相对封闭的社交空间中，缺乏滋生网络暴力的土壤，如今互联网平台的活跃用户往往数以万计，并且逐渐与现实生活和物理空间密不可分，其在网络暴力侵权中的局限性已逐渐显露出来。

### 2.3. “通知 - 删减规则”在网络暴力侵权中的局限性

#### 2.3.1. 从财产权益到人身权益：法益保护范围的局限性

侵权法规则的类型化以被保护法益的类型化为逻辑起点，适格主体、行为模式、过错形态与责任承担等规范要件无一不与法益类型息息相关。在网络版权侵权中，侵权人主要是通过未经许可复制、出版、发行等方式为自己谋取不正当利益，损害版权人的财产权益。而网络暴力侵权表现为“言辞性的精神强制”，其主要侵害的是被害人的人格名誉、身心健康等人身权益。将“通知 - 删减规则”扩张适用于网络暴力侵权中，其最大的理论障碍即在于对固有法益类型基础的突破。《侵权责任法》第 36 条将“通知 - 删减规则”确立为网络侵权的普遍适用规则，彼时我国网络人格权保护制度尚处在起步阶段，通过简单的法律移植贸然扩大适用数字版权法领域规则的立法操作在当时招致了诸多批判。有学者指出，美国的“避风港规则”是网络服务提供商、权利人、网络用户的利益分野在同为财产权利时具有一定的同构性，使得优先保护网络信息产业的整体财产性权利，次优保护个人的数字版权财产权利的价值衡量结果具备

正当性基础，而人身权益的介入将打破这一格局，其与“避风港规则”创立的价值目标定位存在天然的错位[5]。

21世纪的时代精神应该是对人的尊严和自由的保护[6]。诚如《法国民法典》第16条之规定：“法律须保障人的优先性，禁止对尊严的一切侵害，确保自生命伊始对人的尊重。”两大法系几乎都确立了人身利益优先保护的规则，即在财产利益和人身利益发生冲突的人身权益投诉情况下，应当优先保护人身利益。维护好人民的利益是网络暴力治理的首要价值目标，网络用户的人身权益不应向网络信息产业的发展利益让步，盲目的法律移植和扩张适用只会招致“淮橘为枳”的后果。法益保护范围的局限性决定了被动履行义务难以在人身权益保护领域体现出应有的网络治理功能与利益衡平价值。

### 2.3.2. 从个体性侵权到群体性侵权：“点对点”模式的功能性失灵

在“通知-删除规则”的规范路径中，网络平台的被动受理以被侵权用户的主动通知为前提，其立基的现实假设是用户有能力发现、识别侵权信息，并有能力向平台提交侵权的初步证据。在我国的互联网实践中，“通知-删除规则”表现为用户与平台之间的维权投诉机制。以我国规模最大的社交平台“新浪微博”举例，依据《微博投诉操作细则》和《微博人身权益投诉处理流程》，被侵权用户可以通过人身权益投诉入口和一般投诉入口将侵权信息通知站方，站方人员进行初步的核实与确认之后会立即采取删除、屏蔽等措施。一般投诉入口较为快捷，用于处理事实简单清楚的投诉；人身权益投诉入口审核周期较长，用于处理对一般投诉入口结果不满意的情况。其自主投诉流程较为简单方便，在线上投诉页面上平台也设置了相应的指引与提示，然而在既有的网络暴力实践中，该投诉流程对于层出不穷的网络暴力信息的应对基本上是乏力的。如“刘学州网暴事件”中，“新浪微博”仅根据用户举报投诉排查清理了290条泄露个人隐私、挑动矛盾纠纷的违规内容，但在刘学州自杀后，刘学州案的代理律师在固定证据时，在网络上提取了针对刘学州的2000多条性质恶劣的网暴言论，其也仅仅是大量网暴信息中的冰山一角。

网络暴力侵权的危害性表现为攻击规模的海量性，其往往是由成千上万量级的人实施，被侵权人借助于自身力量，基本无法做到全部投诉，难以克服“挂一漏万”的维权困境。并且网络暴力言论侵犯人身权益以侵犯精神性人格利益为主，每一次投诉被侵权人都需要浏览一遍网络暴力言论，对自己的身心健康进行一次又一次的损害。投诉成本无形且高昂，暂时切断与互联网的联系进行自我保护是应对网络暴力的无奈之举，这也同时意味着被侵权人将放弃投诉这一自我救济的渠道，网络暴力信息的传播将再难遏制。“点对点”模式的“通知-删除规则”在大规模网络暴力侵权中几乎是失灵的状态，其功能上的局限性决定了其难以适应网络暴力侵权群体性的特征。

### 2.4. 小结

如前所述，“避风港原则”是上世纪末美国版权业与网信产业比权量力的妥协结果，其证成于独特的时代土壤与利益格局。“通知-删除规则”作为法律移植的产物，扩张使用的本土化操作缺乏理论与功能上的合理性。在网络暴力侵权中，若将其作为认定平台间接侵权过错的核心规范进路，将不合理地缩小限制平台的过错范围，难以实现进一步压实网站平台信息内容管理主体责任的网络暴力治理目标，必须另寻出路。

## 3. 平台安全保障义务的再审视

基于对“通知-删除规则”的反思，有学者提出对《民法典》第1198条安全保障义务条款进行目的性扩张解释，在平台对网络暴力的发生或扩大具有技术能力却未积极履行预防义务或未对受害人尽到合理保护义务时，认定平台应对受害者承担相应的民事侵权责任[7]，并将安全保障义务从平台预防和武装

用户两方面进行内涵构建[8]，这是目前解释论中较为主流的观点。

以安全保障义务条款作为平台间接侵权过错认定的规范进路，其主要目的在于以积极作为义务替代“通知-删除规则”的被动受理义务，促使平台积极参与网络暴力治理，落实平台主体责任。应对《民法典》第1198条规定安全保障义务的初衷在于确立物理场所经营者、管理者制止第三人侵权的积极作为义务。安全保障义务条款以“开启或维系交往空间者对其中的他人负有安全保障义务，应在合理限度内照顾他人权益”作为其法理基础，而在网络暴力侵权中，网络平台作为网络社交空间的创造者、管理者与维护者，其在保障信息安全、规范传播秩序、维护互联网生态等方面具备主体作用。依据《互联网信息服务管理办法》，网络信息内容服务平台应当履行信息内容管理主体责任，应当防范和抵制传播包括网络暴力信息在内的不良信息，应在合理限度内保障平台用户的合法权益。

以安全保障义务条款推动平台主动治理模式的构建值得肯定，实际上也早有学者提出将安全保障义务条款扩张适用于网络平台过错认定的观点[9]。然而安全保障义务条款在一般网络侵权中适用的情形极为罕见，并且从未在网络暴力侵权中予以适用。法理基础上的正当性不等于解释论上的可能性与法律效果上的合理性，平台安全保障义务的证成亟待再审视。

### 3.1. 理论困境：空间属性的解释论障碍

《民法典》第1198条第1款规定：“宾馆、商场、银行、车站、机场、体育场馆、娱乐场所等经营场所、公共场所的经营者、管理者或者群众性活动的组织者，未尽到安全保障义务，造成他人损害的，应当承担侵权责任。”立法者其虽未明示侵权场所的空间属性，但通过列举的方式有意将经营场所与公共场所限制在有形的物理空间之内，但其并非不具备解释论上的空间，网络社交的本身即是巨量级用户参与的群众性活动，将网络平台视为群众性活动的组织者也似乎可以成立。在民法典编纂时，也有人主张应当明确网络服务提供者的安全保障义务，但考虑到盲目扩张安全保障义务的范围会引发更多的社会纠纷等因素，立法机关并未认可该观点。我国的司法实践也对相同的态度，有法院直接指出，将有形物理空间的安全保障义务扩张到无形网络空间，并不利于提高事实与法律规范的契合度，在适用网络侵权的专门规范就能够规则的情况下，没有必要再扩张安全保障义务的适用范围。

有学者认为，安全保障义务源于德国民法的“交易往来安全义务”，其所立基的法理是抽象的风险分配规则，即注意成本应该分配给危险的制造者和保有者以及因特定危险而获得利益者，而非所谓的物理空间，应当打破安全保障义务在适用上的场所限制。依笔者管见，虽然我国的安全保障义务和德国民法的“交易往来安全义务”的内涵大致相同，但是二者在各自的法典中处于不同的规范位阶。“交易往来安全义务”为德国判例学说的产物，其规范基础为《德国民法典》第823条第1款(相当于我国《民法典》的第1165条第1款)，为侵权责任的一般规范；而《民法典》第1198条的安全保障义务条款则位于“责任主体的特殊规定”一章中，体系定位决定了其有限的理论张力与规范功能，不宜过度抽象化其法理。并且第1189条的安全保障义务条款和第1194条到第1197条的网络侵权责任条款均位于“责任主体的特殊规定”，本身就是区分责任主体的规范产物，若将网络平台解释为经营场所、公共场所，将导致二者之间在解释论上的相互龃龉与规制冲突，并造成司法适用上的严重混乱。

### 3.2. 实践障碍：救济路径的重大瑕疵

依据《民法典》第1198条第2款的规定：“因第三人的行为造成他人损害的，由第三人承担侵权责任；经营者、管理者或者组织者未尽到安全保障义务的，承担相应的补充责任。经营者、管理者或者组织者承担补充责任后，可以向第三人追偿。”在网络暴力侵权中，直接侵害行为的实施者是网络用户，平台最多只能被解释为间接侵权人，其过错程度和责任大小无法高于直接侵权人，在《民法典》第1189

条的规范进路下，网络平台将只会承担补充责任，并且其责任承担的顺位在众多直接侵权人之后，如果受害者径行向平台主张权利，平台可以提出先执行前顺位责任的抗辩。相较于“通知-删除规则”的连带责任模式而言，平台补充责任模式将加剧网络暴力侵权的救济困难。由于网络空间的空间虚拟性、主体隐蔽性、群体参与性、模糊性等特征，网络暴力判别尺度、界定治理等方面都存在一定难度，使得侵权法系统运行成本超出运行收益成为常态基于起诉成本等因素考量，补充责任救济的理想状态根本无法得以实现。

补充责任的形态特征主要有两点，一是程序上的次位性，二是责任范围的前位决定性[10]。如前所述，在补充责任模式下，平台享有顺位利益，且劣后于众多网络侵权用户，在网络暴力侵权的诉讼实践中，寄希望于受害者将参与网络暴力的平台用户全部起诉，无异于“痴人说梦”，证据收集与固定的经济成本与时间成本都是难以计量的。“粉发女孩网暴事件”中，受害人花费了4000元，才将15张微博截屏照片、165张百家号截屏照片以及1份刻录的抖音视频光盘加以公证，留作证据。在网络暴力共同诉讼中，受害人要想办法收集侵权人身份信息。依据最高人民法院颁布的《关于审理利用信息网络侵害人身权益民事纠纷案件适用法律若干问题的规定》，原告有权要求平台提供能够确定涉嫌侵权的网络用户的姓名、联系方式、网络地址等信息，然而在网络暴力侵权的诉讼实践中最大的障碍是涉嫌侵权的网络用户的发现和识别。语言暴力的形式复杂多样，且在当下的互联网审核机制下具备一定的隐蔽性，通过“词条”和“关键词”进行检索的方式只确定一部分的侵权用户，“挂一漏万”的诉讼困境难以破解，平台责任难以得到真正落实。

## 4. 普遍审查义务的引入与构建

国家网信办于2024年6月公布的《网络暴力信息治理规定》中，基于网络暴力信息监测预警机制、网络暴力信息处置和受害者保护机制的分阶段展开，尝试从技术规范的层面构建平台的普遍审查义务体系。已有学者提出，应以此为契机，通过侵权法规则与技术规范的有效衔接，重塑“知道规则”中的义务标准，并让“知道规则”替代“通知-删除规则”在解释论上的核心地位[11]。相较于安全保障义务条款，

“知道规则”作为连带责任条款和既有的互联网侵权责任专款，其并不存在上述的理论障碍与救济路径的重大瑕疵，然而将普遍审查义务引入“应当知道”的内涵解释中，其正当性亟待进一步证成。如何将技术标准转化为侵权法中的义务标准，也是构建普遍审查义务的难点与核心。

### 4.1. 引入普遍审查义务的理性基础与规范基础

将普遍审查义务引入“知道规则”，最大的理论障碍在于正当性的证成上。网络服务提供者不负有普遍审查义务已得到我国学者的普遍承认，我国学者对该观点的认识主要塑成于《侵权责任法》的立法过程中，体现在第36条第3款能否采用“知道或应当知道”这一立法表达的争论上。典型观点如杨立新教授指出的，网络平台不同于传统媒体，海量信息造成事先审查的客观不能，基于“技术中立原则”，网络服务提供者不负有事先审查义务[12]。

“技术中立原则”也被称作“实质性非侵权用途原则”或“普通商品原则”，其肇始于1984年的“索尼案”，案件的焦点在于索尼公司是否应当为他人使用索尼生产的录像机进行版权侵权行为而承担间接侵权责任，美国最高法院认为，单纯的技术提供行为缺乏承担侵权责任的正当性，如果“产品可能被广泛用于合法的、不受争议的用途”即“能够具有实质性的非侵权用途”，即使制造商和销售商知道其设备有可能被用于侵权，也不能推定其具有帮助他人侵权的过错并构成间接侵权[13]。美国最高法院在“索尼案”中确立了“技术中立原则”后，又将其广泛适用于网络版权侵权纠纷中作为网络服务提供者的抗辩事由，《千年数字版权法》中的“避风港原则”亦是“技术中立原则”的延伸。

但是应当注意的是，《千禧年数字版权法(DMCA)》颁布于1998年，彼时的网络世界尚处于Web1.0时代，网络空间所聚合的对象是巨量、芜杂的网络信息<sup>[14]</sup>，网络服务提供者仅提供技术与通道支持，其角色仍以被动性、工具性和中立性为特征，并没有主动参与网络信息内容的生成与传播，网络侵权的危险源的开启与控制与网络服务提供者无关。然而二十年过去，人类已然进入了Web3.0时代，现实世界和虚拟世界的鸿沟正在不断缩小，平台的定位也在悄然发生改变，其往往兼具网络服务提供者和信息内容提供者的双重身份，并主动参与到网络内容的生成与传播的过程中，以“流量推广”、“个性化推荐”等方式积极推动网络平台上信息传递的频繁进行和规模递增，对危险源的产生存在着一定的原因为。

“技术中立原则”确立之初还有着鼓励和保护技术创新的考量<sup>[15]</sup>，诚如德国社会学家乌尔里希·贝克所言：“专业化程度越高，科技活动的副作用在范围、数量和不可计算的程度等方面也会变得越大。”<sup>[16]</sup>在信息技术高速发展的时代，网络服务提供者往往无法预知技术的缺陷及其可能带来的危险，若使其因此负担过重的侵权责任，会对信息技术的发展创新造成不利影响。然而对于网络社交平台而言，其对技术创新的需求和贡献都不算高，其技术发展主要是基于对即时通讯技术的优化，技术经验相对成熟，缺陷和风险都是有迹可循，并且可以预见到的。基于“技术中立原则”过度保护网络社交平台的产业利益，不仅无益于网信产业的良性发展，更会导致诸如AI识别技术等反网暴技术的停滞不前。并且近年来信息技术日新月异，网信产业的蓬勃发展，网络服务提供者不再是单纯的通道提供者，通过较大量级的浏览量和点击量来吸引广告商投放广告以获得巨大的经济利益，也使得其担责能力得到了巨大的提高。新的利益格局极大地动摇了作为网络侵权中技术中立原则的正当性，网络服务提供者的利益与负担不均衡亟待进一步调整。

网络暴力侵权相较于其他网络侵权的特殊性，在于其所侵害的主要法益是公民的健康权乃至生命权，具有法益侵害上的严重性和紧迫性，对网络服务提供者的注意义务的要求显然要高于网络著作权侵权和其他网络人格权侵权。善良管理人之安全保障义务具有中立性和被动性的色彩，无法满足我国当下网络暴力治理的现实需要。并且网络暴力具有“强扩散性”，绝大多数网络暴力侵权都表现为多数人侵权的形态，“热点事件”常常“一石激起千层浪”，随着网络暴力行为的持续会有越来越多的施暴者加入其中。删除、屏蔽和断开链接等有限的技术手段终究无法和大规模的网络暴力行动抗衡，“打地鼠效应”和“跨平台传播”使得网络暴力的规模不断扩大，最终导致时间上、范围上的完全失控。德国社会学家诺依曼提出“沉默的螺旋理论”解释该类现象的成因，“如果看到赞同的观点有多数人认可，就会果断广泛传播，而对于少数人支持的观点，则为了避免被孤立而选择沉默。依次循环，就会形成不同观点之间此消彼长的螺旋过程”<sup>[17]</sup>，在网络空间中，网络暴力言论可以借助大众传播媒介在短时间内实现“自下而上”的扩散，并迅速形成优势意见，其他意见则因为个体为了避免被孤立而趋于沉默，优势意见最终形成压倒性的舆论优势。在网络暴力形成的初期往往只有零星的网络负面评论，当评论的数量不断累计，“意见气候”的质变也随之产生，并借由“沉默的螺旋”不断实现规模的扩张。网络暴力治理需要平台在“意见气候”形成之前及时行动，赋予网络服务提供者以普遍审查义务，其主要目的在于未雨绸缪，尽可能避免网络暴力侵权的发生。

最高人民法院于颁布的《关于审理侵害信息网络传播权民事纠纷案件适用法律若干问题的规定》(以下简称《网络传播权侵权纠纷司法解释》)中指出，“网络服务提供者未对网络用户侵害信息网络传播权的行为主动进行审查的，人民法院不应据此认定其具有过错”，有学者以此为根据得出了平台的普遍审查义务不具备规范基础的结论<sup>[18]</sup>。应当注意的是，尽管《网络传播权侵权纠纷司法解释》明确否定了平台的普遍审查义务，但是《网络人身侵权纠纷司法解释》中却并未如此规定，并在第9条将“网络服务提供者采取预防侵权措施的技术可能性及其是否采取了相应的合理措施”列入“知道”的判断标准之中。最高人民法院在制定《网络人身侵权纠纷司法解释》显然考虑到了人身权益相对于财产权益而言更高的

保护位阶，所以并未将普遍审查义务彻底排除在外。虽然《网络人身侵权纠纷司法解释》尚有针对性不强等诸多不尽人意之处，但是以其为规范基础不断完善义务标准，构建技术性规范与侵权法规则相互沟通的桥梁，仍然是当下最有效、最直接的改进方案。

## 4.2. “应当知道”的内涵重塑：比例原则下普遍审查义务的构建

普遍审查义务最大的实践障碍在于高昂的技术成本以及可能对公民造成不当限制，本文认为，可引入比例原则作为利益平衡的工具：审查义务的技术实现路径需要满足均衡性的要求，平台承担的技术成本应与欲达致的网络暴力治理目标相称；审查的范围与程度需要满足必要性的要求，避免对公民造成不当限制。下文将详细展开论述。

### 4.2.1. 均衡性考量下普遍审查义务的技术实现路径

对网络信息内容的审查存在人工审查与技术审查两种，人工审查仅适合在个案中进行，对海量信息的普遍审查则依赖于大规模机器审核和网络内容过滤技术的应用。引入普遍审查义务最大的争议，即是其易造成的网络服务提供者较大的技术成本负担，对部分规模较小、盈利能力不强的平台而言，甚至可能产生即便承担侵权责任，也不愿意引入并维护相关技术的窘境。比例原则要求手段与目的之间应保持均衡，即手段的成本与其所增进的利益是成比例的[19]。普遍审查义务的技术实现也应满足均衡性的要求，本文认为，可借鉴传播学中“沉默的螺旋”理论，平台的普遍审查义务分阶段展开，将有限的技术成本投入到“优势意见形成后”的关键环节中去。

“沉默的螺旋”理论常被用来解释网络暴力的形成过程，即如果看到赞同的观点有多数人认可，就会果断广泛传播，而对于少数人支持的观点，则为了避免被孤立而选择沉默。依次循环，就会形成不同观点之间此消彼长的螺旋过程。网络暴力的发展进程是螺旋式的，网络暴力治理也是分阶段的，在“优势意见形成前”、“优势意见形成后”和“大规模网络暴力形成后”的不同阶段，权益保护的紧迫程度与对审查技术的需要也有所不同。

在“气候意见”形成前，即只存在零星的网络负面评论时，平台无需担负过重的审查义务。一是因为少量的网络负面评论尚未形成规模，影响力并不大，不太可能导致越来越多人加入的“破窗效应”的发生。二是因为“通知-删除规则”尚有用武之地，被侵权人通过线上投诉即可实现自力救济。故而此阶段的审查义务标准不宜设定过高，平台能够将违法信息和明显的不良信息过滤即可。该阶段属于监测预警阶段，宜引入《网络暴力信息治理规定》第三章“预防预警”的规定，将网络暴力信息预警模型、信息特征库和典型案例样本库是否健全作为是否履行了审查义务的客观标准。预警模型和信息样本库是网络暴力治理的技术基础，主要通过发展关键词拦截策略、行为识别模型、标记重点用户等方式，及时遏制“气候意见”的形成，其合格与否不在于完美周延地精准识别所有的网络暴力信息、管理海量相关账号，而在于能否精准监测、识别网络暴力信息高发的内容场景。

在“气候意见”形成后，即网络暴力形成一定规模，多数意见开始扩张之时，平台需要参考网络暴力的舆情周期(一般是一周左右)，承担启动一键防护的协助义务、合理期限内的持续人工主动审查义务、以及具体事件中网暴行为识别模型的构建义务，避免大规模网络暴力事件的发生。我国目前多数社交平台都在应用内设置了“一键防护”功能，只要开启即可在一段时间内屏蔽掉所有的私信和评论。但是鉴于多数用户对这一新功能并不熟悉，平台有义务在“气候意见”形成后协助启动防护措施。并且，平台也有义务建立具体事件中网暴行为的识别模型，参考具体网络暴力事件的主要内容，信息语境和舆论导向等内容，为其设定具体的信息过滤的识别标准，通过不断充实具体事件中的典型案例样本库以便最大程度过滤网络暴力信息。此阶段的审查工作不宜只局限于机器审查，需要定期的进行词条下或关键词下的人工审核，尽最大程度识别隐蔽性较强的网络暴力信息。

在大规模网络暴力形成后，即发生跨平台传播等单以平台采取措施无法遏制网络暴力蔓延的情况时，平台的审查对象将不限于本平台的网络暴力信息，跨平台的网络暴力信息也应当得到注意。一是应当将平台内以及跨平台传播的情况，及时通知网信部门，履行好报告义务，做好信息共享、会商通报、取证调证等工作，以便网信监管部门进一步治理和救济工作的展开。二是要积极和其他平台协作治理网络暴力，与其他分享相关数据和具体事件中的行为识别模式，严防死守网络暴力的进一步蔓延。

从实践的经验来看，大规模的网络暴力主要产生于市场估值体量巨大、用户规模巨大的超大型或大型互联网平台，具备较强的资源技术能力来履行普遍审查义务[20]。当小型或中型平台往往不具备落实并维护相关技术的经济能力，因此不宜对它们苛以同样的普遍审查义务。本文认为参考《欧盟数字服务法》关于分层监管的规定，中小型平台可以选择有限度地承担普遍审查义务，或选择履行行业协会更为宽松的自律标准，避免硬性监管要求超出了平台的实际需要。

#### 4.2.2. 必要性考量下普遍审查义务的范围

普遍审查义务的负担同时也意味着平台审查权力的强化，网络暴力治理下公民权利的维护也同样值得重视。比例原则下的均衡性原则要求对基本权利的干预与其所追求的目的之间必须要相称，普遍审查义务对网络暴力治理增进的利益必须与对言论自由的限制相称，这就需要对普遍审查义务的范围进行一定程度的控制。

首先，平台应当将普遍审查义务的目标限定在网络暴力信息，将正常的监督与批评言论排除在外。这需要明确细化网络暴力信息标准，并在技术上增强识别的准确性。依据《网络暴力信息治理规定》，网络暴力信息是指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道德绑架、贬低歧视、恶意揣测等违法和不良信息。对于侮辱、诽谤和侵犯隐私等明显侵害人格权益的“违法信息”进行审查自不必言，其以是否违反法律规范为判断标准，具备语言形式上的明显性与确定性，通过既有的关键词过滤技术即可得到识别，关键问题在于“不良信息”范围的界定上。

国家网信办于2022年发布的《关于切实加强网络暴力治理的通知》中将“不良信息”界定为“不友善信息”，但是在真实的网络舆论环境中，监督和批评往往伴随着网民的负面情绪，在语言形式表现出“不友善”的特征，“不友善信息”显然过度扩张了“不良信息”的限定范围。有学者提出应参考网络暴力发生的特有场域，对“不良信息”进行具体认定。对此，笔者持赞同态度。网络语言具有创新性、流行性、局域性三个特征，语境差广泛地存在于网络语言中[21]。同样的文字在不同的社交场景中可能呈现出完全相反的意思，在不同的平台文化空间中会生出不同网络暴力“黑话”和侮辱表情包等“梗文化”[22]。如在武汉“碾压案”妈妈自杀事件中，小学生被碾压身亡，母亲只是因为带妆出镜，即遭受到诸如“打扮是用了心的”、“小孩很可怜，但家长总买热搜炒作就不合适了”之类言论的网络暴力，此类言论看似不具备侵害性，但是放在具体语境下，对受害者的打击并不亚于直白的侮辱。

不良信息与正常的批评言论的界分在技术上并非无法实现，自然语言处理技术(NLP)能够做到“不良信息”的有效识别，通过情绪解析、文本观点提取、文本相似度判断和词法分析等手段，使计算机更接近于人类对语言的理解。如“微博智搜”的大众情感分析系统，通过系统随机抓取热点话题中的内容作为样本，然后通过AI算法分析用户情绪，并总结综合观点，以服务于平台舆情治理，其在网络暴力治理中将大有作为[23]。但是自然语言处理技术(NLP)作为较为前沿的技术，在我国尚未推广普及，并且机器学习需要大量数据的素材积累，和学习模型的不断调试，其技术成本较高，并非所有平台都有能力负担。

通过不完全列举或概括总结的方式，从理论上对“不良信息”的范围予以界定绝非难事，但在网暴信息识别技术尚不成熟的当下，将其全部列入普遍审查的义务范围，不仅会使平台的网暴治理负担过重，还有可能会导致不分平台出于规避间接侵权风险的考量，不当地扩大算法的识别范围，限制网络言论自

由。目前我国的各大主流社交平台都对不良信息的认定问题制定了相应规范，如“微博社区公约”、“抖音社区自律公约”、“哔哩哔哩不友善行为”和“知乎社区规范”等，均依据平台的言论环境和用户肖像和内容社区氛围制定了不同的类型标准，具备实际应用中较强的针对性，并且通过既有的信息发布审核技术即可得到实现。依笔者管见，宜将不良信息的认定标准下放到各平台。相较于网信部门等公职机构，平台更了解本社群的情况，且能对新型的不良信息及时予以回应。平台宜从道德标准、社区标准和公序良俗标准三个角度出发，以类型化的列举方式划定不良信息的范围。同时为了避免平台推卸责任，而不当限缩不良信息范围的现象发生，应由国家网信部门、行业自律组织和其他第三方共同对其自定标准的合理性予以核查。宜从采取预防侵权措施的技术可能性出发，兼顾行业平均水平、平台经济实力和用户数量，将是否满足能够及时有效治理网络暴力的实质性标准作为合格与否的判断依据。

### 4.3. 公私法衔接视角下普遍审查义务的落实

在平台的过错认定中引入普遍审查义务，能够为网络暴力的受害人提供更为充分的私法保护。然而普遍审查义务的落实绝不仅仅是民事实体法的任务，也需要民事实体法与行政法规范和程序法规范的衔接。

首先，我国现阶段的网络暴力治理正处于法治化转型及立法体系建构的过程中，民事实体法与行政法规范的互动应当得到充分的重视。理论上一般认为，违反行政法规范并不能直接认定过错，但行政法规范所提供的行为义务标准能够为过错判定时的注意义务标准提供参考<sup>[24]</sup>。本文认为，能够作为参考标准的行政法规必须是以用户为主体的保护性规范，并且不能根据保护性规范的遵守反推平台不存在过错。第一，保护性规范的目标是维护民事主体权益，因此存在与侵权法接轨之可能，非保护性规范则只是为了维护公共利益。无法作为过错的判断标准。《网络暴力信息治理规定》即是典型的保护性规范，其对平台要求的技术标准可作为是否违反普遍审查义务的判断标准。第二，由于行政法规范往往提供的是社会生活中的较低行为标准，且具有一定的滞后性，因此平台不能以没有违反行政法规作为过错的抗辩理由。

其次，普遍审查义务的引入虽然推动了平台义务标准的高阶化，但是网信前沿技术的融入也使得受害者对有关证据的收集和证明变得更加困难，本文认为有必要在程序法上引入民事检察公益诉讼。早在2023年，“两高一部”于2023年联合印发了《关于依法惩治网络暴力违法犯罪的指导意见》已经提出了将民事检察公益诉讼的受案范畴扩展至网络暴力侵权的设想。在平台未履行普遍审查义务，致使网络暴力信息大量传播造成足以威胁社会公共利益的严重后果时，人民检察院可以依法提起公益诉讼。公益诉讼制度能够有效弥补受害人和网络平台法律地位上的实质不平等，通过检察机关和其他网络监管部门的跨部门合作，能够有效地弥补了用户证据收集与证明上的不足。2025年10月，《检察公益诉讼法(草案)》公布，可借此立法契机将网络暴力正式纳入民事检察公益诉讼的受案范围中，以期为普遍审查义务的落实提供更有力的组织与程序保障。

## 5. 结语

我国的网络暴力治理正处于法治化转型的关键时期，考虑到网络暴力侵权后果的严重性与易传播性，对平台的注意义务要求也应从被动的事后受理向主动的普遍审查转变。在比例原则下，普遍审查义务的构建需要满足均衡性与必要性的要求，以避免过高的成本负担或对言论自由的不当限制。普遍审查义务的落实既需要侵权法规则与行政法规范形成合力，也需要民事检察公益诉讼提供程序与组织保障。本文聚焦于责任认定阶段，力图在解释论框架下为平台的过错认定提供方案，但责任承担阶段中平台与网暴者的责任范围划分、人格权保护禁令的适用，以及损害赔偿责任与预防性侵权责任的衔接等问题同样值

得重视，有待进一步研究。

## 参考文献

- [1] 张新宝,任鸿雁.互联网上的侵权责任:《侵权责任法》第36条解读[J].中国人民大学学报,2010,24(4):24-25.
- [2] 赵泽睿.平台革命引发的美国版权责任变革及经验分析[J].电子知识产权,2020(12):34.
- [3] 刘晋名,艾围利.“避风港规则”的法律适用困境及消解路径[J].南京社会科学,2020(8):95-99,116.
- [4] United States Copyright Office (2020) Section 512 of Title 17 A Report of the Register of Copyrights.
- [5] 陈昶屹.“避风港规则”扩张适用网络人格权保护之困境与消解——兼论侵权责任法第三十六条之完善[J].人民司法,2012(1):80-86.
- [6] 王利明,周友军.我国《民法总则》的成功与不足[J].比较法研究,2017(4):1-15.
- [7] 石佳友.网络暴力治理中的平台责任[J].法律科学(西北政法大学学报),2023,41(6):14-23.
- [8] 刘文杰.网络暴力中个体维权困境与平台安全保障义务[J].国家检察官学院学报,2023,31(5):39-57.
- [9] 刘文杰.网络服务提供者的安全保障义务[J].中外法学,2012,24(2):395.
- [10] 刘海安.侵权补充责任类型的反思与重定[J].政治与法律,2012(2):121-131.
- [11] 刘金瑞.网络暴力侵权法规制路径的完善[J].政法论坛,2024,42(3):72-76.
- [12] 杨立新.《侵权责任法》规定的网络侵权责任的理解与解释[J].国家检察官学院学报,2010,18(2):3-10.
- [13] 王迁.“索尼案”二十年祭——回顾、反思与启示[J].科技与法律,2004(4):59-68.
- [14] 段寿建,邓有林.Web技术发展综述与展望[J].计算机时代,2013(3):8-10.
- [15] 何培育,刘梦雪.技术中立原则在信息网络传播权保护领域的适用[J].重庆邮电大学学报(社会科学版),2017,29(3):42-50.
- [16] [德]乌尔里希·贝克.风险社会:新的现代性之路[M].南京:译林出版社,2018.
- [17] [德]伊丽莎白·诺艾尔·诺依曼.沉默的螺旋:舆论——我们的社会皮肤[M].北京:北京大学出版社,2013.
- [18] 冯术杰.论网络服务提供者间接侵权责任的过错形态[J].中国法学,2016(4):179-197.
- [19] 刘权.均衡性原则的具体化[J].法学家,2017(2):17-30,177-178.
- [20] 王燃.论网络暴力的平台技术治理[J].法律科学(西北政法大学学报),2024,42(2):121-134.
- [21] 黄涛,赵涵,叶静雯.从热梗看网络语言的语境差[J].海峡人文学刊,2021,1(4):57-64.
- [22] 印波,庄新宇.社交平台反网络暴力专项合规机制的有效构建[J].江西社会科学,2024,44(1):93-103.
- [23] 奚雪峰,周国栋.面向自然语言处理的深度学习研究[J].自动化学报,2016,42(10):1445-1465.
- [24] 朱虎.规制法与侵权法[M].北京:中国人民大学出版社,2018.