

网络违法行为数字平台治理的法律规制研究

于海文, 沈悠妍, 周诗盈

浙江农林大学文法学院, 浙江 杭州

收稿日期: 2026年7月3日; 录用日期: 2026年8月19日; 发布日期: 2026年8月27日

摘要

伴随着信息技术的发展, 网络违法行为借助数字平台不断滋生和蔓延, 对国家安全、社会秩序和公民权益构成了严重威胁。文章对网络违法行为类型化、平台治理法理基础的梳理, 对立法现状进行整理, 分析了数字平台对网络违法行为治理面临的现实困境, 主要包括以下几点: 技术迭代与网络违法行为治理滞后的矛盾, 平台预防性义务和监管义务虚化问题以及法律对违法行为认定不一、对平台权力限制规制问题。在完善数字平台对网络违法行为的治理方面, 文章重点提出完善算法设计强化平台智能监管、前置平台审查责任、加强对平台的监督审查机制、明确网络违法行为认定标准、构建数字平台滥用治理权限的防范机制等方式, 推动平台遏制违法行为, 营造健康、良性的网络环境, 守护国家网络安全。

关键词

网络违法行为, 数字平台治理, 平台责任, 平台权力

Research on Legal Regulation of Digital Platform Governance for Online Illegal Acts

Haiwen Yu, Youyan Shen, Shiying Zhou

College of Humanities and Law, Zhejiang A&F University, Hangzhou Zhejiang

Received: July 3, 2026; accepted: August 19, 2026; published: August 27, 2026

Abstract

With the development of information technology, cyber illegal activities have continuously proliferated and spread through digital platforms, posing a serious threat to national, social order, and citizens' rights and interests. This article categorizes cyber illegal activities, reviews the jurisprudential basis of platform governance, and summarizes the current legislative status. This article

文章引用: 于海文, 沈悠妍, 周诗盈. 网络违法行为数字平台治理的法律规制研究[J]. 社会科学前沿, 2026, 15(8): 754-762. DOI: 10.12677/ass.2026.158712

analyzes the practical dilemmas faced by digital platforms in governing cyber illegal activities, which mainly include the following points: the contradiction between technological iteration and the lag in governing cyber illegal activities; the of platforms' preventive and regulatory obligations; and the inconsistent legal determination of illegal activities and the regulation of platform power restrictions. To improve the governance of cyber illegal activities by digital platforms, the article focuses on proposing measures such as improving algorithm design to strengthen intelligent platform supervision, advancing platform review responsibilities, strengthening the supervision and review mechanism for platforms, clarifying the determination standards for cyber illegal activities, and constructing a prevention mechanism against the abuse of governance authority by digital platforms. These measures aim to promote platforms to curb illegal activities, create a healthy and positive online environment, and safeguard national cybersecurity.

Keywords

Cyber Illegal Activities, Digital Platform Governance, Platform Responsibility, Platform Power

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. 网络违法行为数字平台治理的基础理论

1.1. 网络违法行为的范畴与类型化

随着互联网行业的发展和个人通讯设备的普及,网络空间打破了时空的限制,成为了人类活动的“第三空间”。但是互联网的虚拟性、匿名性、高效性、互动性等,使其成为众多违法行为的实施工具或发生场所。网络违法行为即利用信息网络及相关技术,违反国家法律法规,侵害国家、社会公共利益或他人的合法权益的行为。

对网络违法行为进行梳理,基于其行为方式进行划分,可将其分为直接实施型网络违法行为和帮助型网络违法行为。

1.1.1. 直接型网络违法行为

直接型网络违法行为是指行为人主动利用网络技术实施违法行为,直接侵害国家、社会公共利益或他人的合法权益。这类行为是网络违法活动的核心与源头,其行为主体通常是违法利益的直接获取者或危害结果的直接造成者。如电信诈骗、散布虚假信息 etc.

1.1.2. 帮助型违法行为

间接型网络违法行为是指行为人并未直接实施侵害法益的核心行为,但其行为为直接型网络违法行为的发生、实施或危害后果的扩大提供了关键的支持、帮助、或创造了风险环境。这类行为主体可能是直接违法行为的共犯,也可能是具有监管责任的平台方等,其行为方式往往更具隐蔽性。如提供网络账号出租服务帮助网络赌博,为违法网站进行推广等。

1.2. 平台参与网络违法治理的法理基础

1.2.1. 获益与风险相一致原则

伴随着互联网技术的不断发展,数字平台在日常生活中变得愈加重要,越来越多的违法行为与数字平台相关联。在直接型网络违法行为中数字平台往往是发生场所,如在电信诈骗中行为人与受骗者往往

是在数字平台中建立初步联系和信任,DDoS 攻击中目标也往往是数字平台;帮助型违法行为中数字平台往往作为实施工具或风险环境出现,如网络传播淫秽物品往往是数字平台不能完全履行审查义务而提供了犯罪环境。数字平台因其技术能力与资源垄断地位,在网络空间中具有巨大经济利益与社会影响力,也需要承担相应的社会责任对网络违法行为进行治理和预防。学者周濂认为权力与责任必须对称[1]。平台享有诸如算法推荐、规则制定等红利,就必须承担与之匹配的违法治理和风险控制义务。根据马怀德、王晓淑的研究,平台企业通过数据聚合与流量分发形成了一种特殊的权力[2],这种权力本质上源于国家对数字平台的授权与用户权利的让渡。陈荣昌进一步指出,平台对数据的实际控制权使其成为重要且必须的治理主体,只有将网络平台纳入治理现代化语境,强调治理主体多元互动、信息共享,才能更好平衡多元利益,促进数据领域有序发展、行业创新与权利保障之间的动态均衡[3]。因此,基于权力责任相对应的法理逻辑,获益与风险相一致原则构成平台参与治理的核心要求。

1.2.2. 平台治理网络违法行为的正当性边界

让平台参与对网络违法行为的治理当中是发展的必然趋势也是现实需要,但由于平台私有性带来的自利性,平台权力行使必须受严格限制,否则可能催生出不可控的数字权力支配机制。从平台与国家关系来看平台应禁止权力越位,从平台与私人关系来说平台应坚守数字正义理念。

1) 禁止权力越位

网络空间是人类活动的新空间,网络空间是国家主权行使的重要领域。国家对本国的网络空间具有管辖权,“管辖权是指主权国家对于网络疆域内的一切人、事、物以及网络疆域外的本国人实行管辖的权力,有权按照自己的情况治理网络空间,并制定法律和规则。”[4]平台可协助识别违法信息,按照平台的规定进行处理,但司法裁决与刑罚执行权必须由国家垄断。学者郝英兵也认为私人主体无权替代公权力实施惩戒[5]。网络平台作为网络空间的一部分,仍需接受国家的监督。

2) 坚守数字正义理念

数字正义理念追求数据资源得到合理分配、数字权利得到充分配置、算法决策公开透明化。数字正义理念要求平台尽力保障“消解不同数据主体之间数据占有、使用的不平等,实现参与平等和结果正义”[6]。要求保障平台用户的权利,尊重用户的隐私权,发布、删除和更改信息的权利等。算法决策流程的公开需要符合普通用户的常识判断和理解能力,必要时应以通俗易懂的方式向被决策对象进行解释,即“用可理解的术语向人类表达”[7]。在数字平台自身的发展完善过程中,需要坚持合理配置权利义务,尊重用户的信息与权利,规避“黑箱操作”。

2. 网络违法行为数字平台治理的现状与困境

2.1. 网络违法行为数字平台治理的现状

2.1.1. 立法现状考察

目前我国已构建起较为完善的网络违法行为数字平台治理法律体系,其覆盖数字平台责任履行等各领域,为规范平台治理提供了坚实保障。立法明确平台全链条监督审查责任,要求建立违法行为处理机制。

就总体而言,立法仍存在算法治理领域规则滞后、部分条款可操作性低的问题。未来立法应注重加强对数字平台的分类分级监管,推动数字平台适配经济与技术发展需求。

2.1.2. 实践案例分析

1) 北京爱奇艺科技有限公司与吴某网络服务合同纠纷案

本案中,爱奇艺公司对享有“热剧抢先看”VIP 会员权益的协议,增加了“付费超前点播”条款的行为,引发了争议。吴某认为爱奇艺公司推出的“付费超前点播”模式,损害其会员权益;但爱奇艺公司却

持有相反意见。最终，二审法院经审理认为，“涉案VIP会员协议”构成格式条款，其导言第二款相关内容无效；爱奇艺公司增加的“付费超前点播”条款对吴某不发生法律效力，其行为构成违约，应当承担违约责任¹。

通过本案，可以看出，平台意图通过格式条款的形式排除对法律强制性条款的适用，属于对消费者不公平的规定，应当认定为无效。因此平台在为消费者提供服务时，应尽到安全保障义务和合理提示义务，否则可能会承担相应责任。

2) 李某侵犯公民个人信息“人脸识别”案

本案中，被告人李某制作一款具有非法窃取安装者相册照片的手机黑客软件，窃取安装者相册照片，其中部分照片含有公民个人信息。之后被告人李某在暗网论坛购买“社工库资料”并转存于网盘，并将网盘链接分享至QQ群，提供给群成员免费下载²。法院经审理认为，被告人李某的行为已构成侵犯公民个人信息罪。同时其非法获取并向他人提供公民个人信息的侵权行为，侵害了众多公民个人信息安全，损害社会公共利益，应当承担相应的民事责任。

通过本案，可以看出我国法律严厉打击侵犯公民个人信息的行为。未来法律仍会重点关注不法分子利用网络平台实施违法犯罪行为，维护社会公平正义。

2.2. 网络违法行为数字平台治理的困境

2.2.1. 法律体系碎片化

1) 网络违法行为认定标准不一

若想发挥数字平台对于网络违法行为的治理效能，便要明确违法行为的认定标准。但是不同法律对于网络违法行为的认定欠缺统一标准，这也为平台治理行为的实施带来难题。

现行法律对于网络违法行为的认定多为原则性条款，例如有关条文对于“知道或应当知道”的表述，在不同语境下即可作不同理解。网络违法行为同时受到公法与私法的双重调整，但是二者在法理逻辑和行为认定上具有较大差异，这将会造成规范适用和证明标准的混乱。如今网络技术迭代更新速度较快，立法相较于新型违法行为的出现具有一定的滞后性，这也导致数字新领域中行为的定性缺乏明确依据。

2) 缺少对平台滥用权力的限制性规定

平台自身权力受制于平台经济发展特点而偏离理想的预定轨道，目前来看平台经营者滥用权力侵害用户权益的行为在其规则制定和措施实施的各环节均有发生。在规则制定层面，数字平台忽视对用户主体知情权的保障。平台往往会凭借自身优势地位在平台服务协议的内容上做出减免自身风险责任、加重对方义务的决定^[8]。在措施实施层面，由于行为的管制决策更多取决于平台的主观意志，因此平台经营者的权力行使会存在较大的弹性空间。虽然法律要求平台经营者应审查申请进入平台的商户资格，但审查标准未能明确，这间接增加平台滥用权力的可能性。

平台经营者通过汇聚海量用户的私权利与社会的公权力，不断强化自我管制能力，并推动自身权力的持续扩张与延伸。平台所具备的双边市场属性，决定了其与用户之间在实质地位上的不对等。这种力量失衡直接瓦解了法律所追求的“主体给付平衡”——即通过形式平等保障双方权利义务的对等性与交换合理性的基础^[9]。尽管现行法律法规已在很大程度上限制了平台自身权利的滥用，但当平台经营者对用户形成事实上的“控制与支配”关系时，双方地位的显失平等不可避免，这也进一步放大了数字平台滥用权力的风险。例如在前文所述“北京爱奇艺科技有限公司与吴某网络服务合同纠纷案”中，爱奇艺平台滥用自身权力和市场地位，侵害用户合法权益，因此有必要在法律层面进一步细化平台行使自身权

¹(2020)京04民终359号。

²(2021)沪0120刑初828号。

力的条件与限度要求,以有效规制其滥用行为。

2.2.2. 技术迭代与网络违法行为治理的矛盾

1) 网络算法机制存在缺陷

随着人工智能的发展,网络算法机制在数字平台中得以被广泛应用。虽然算法机制在信息传递、金融算力运行等方面发挥着积极作用,但是机制自身的缺陷也为个人信息保护和网络安全带来隐患。针对数字平台而言,该隐患主要表现为算法机制与用户个体间信息获取差异化,即算法依据个人浏览记录等特征为用户个体创造微信息环境,随着算法与个体互动的深入该环境会不断同质化进而降低个体对多元化信息的偶遇能力,通常该种情形也被称为信息茧房[10]。这将会导致平台用户的视野受制于特定的信息圈层,丧失全方位了解网络的机会,甚至会盲从不良信息,引发网络暴力的风险。部分用户为消除此种风险会“反向驯化”网络算法机制,但最终依然有陷入主体性消解的风险,加剧主动探索能力的退化[11]。

在互联网时代的大背景下,数据作为信息交流的媒介发挥着不可替代的作用。若此时数字平台经营者凭借其市场支配地位采取数据垄断行为,如达成以算法合谋为主要形式的垄断协议、利用数据资源滥用市场支配地位、数据驱动型经营者集中等数据垄断形式,这将会给数字经济和网络媒体发展带来难以估量的损失[12]。

综合以上算法机制的缺陷表现,究其原因是网络算法机制自身的运行风险,即“算法黑箱”。算法黑箱是指人类与人工智能交互,数据输入至结果输出过程中所不能或无须知悉的存在[13]。鉴于算法的隐蔽性,算法黑箱会使数字平台用户难以察觉算法的运行机制和操作原理,从而导致侵犯泄露个人隐私的不利后果,造成用户对数字平台的信任危机,破坏信息网络秩序。

2) 平台治理行为滞后

当前法律法规与大部分学者对于发挥数字平台在治理网络违法行为中的主体作用持有支持赞同的态度,认为应促进平台自治模式的形成,使其充当“类政府”的监管角色。但根据实际治理情况来看,成效往往并不显著,其中的一个重要表现便是平台治理行为滞后。这意味着网络违法行为出现时通常不能在第一时间被平台发现并加以管制,只有造成严重的危害结果才会引起平台主体注意进而采取删帖封号、加以警告等处罚措施,甚至需要公安部门通知,但此时数字平台已经丧失其先行监管的功能。

造成平台治理行为滞后的原因通常有以下几点。首先,平台私权力过度扩张、公权力介入不足。平台凭借技术和数据优势,过度自我赋权,其制定的协议实质性地影响用户的使用和救济权利。目前诉求渠道的阻塞和有效信息的遮蔽时常发生,造成流言以及小道消息兴起,导致诉求转译难度较大,而公权力机关却未能采取行之有效的措施约束其权力[14]。其次,违法行为渐趋隐蔽、平台监管规则滞后。网络违法行为已呈现技术化特征,利用多种新型手段规避平台和法律追踪,其往往散布在海量数据之间,使得传统的监管模式难以有效锁定行为人。例如在前文所述“李某侵犯公民个人信息‘人脸识别’案”中,李某利用各种技术手段逃避平台管控,在网络散布个人信息,给公民和社会带来严重危害,这充分体现了平台监管的不足和滞后。因此,数字平台应注重变革治理方式,从而有效地防范网络违法行为出现。

2.2.3. 数字平台治理责任虚化

1) 忽视平台预防性义务的履行

数字平台作为网络市场的经营者和获益者,应当承担起预防网络违法行为发生的义务,但实际上数字平台往往都是在事后才会进行纠错与规制。

数字平台忽视其预防性义务的履行,主要呈现为以下几点。首先,数字平台以企业获利为首要目标而致使预防性平台违法行为治理的正义性表达缺失[15]。多数平台在履行预防性义务的过程中偏重效率,未把公平放在首位,凸显解决纠纷的效率优先,弱化评审结论的公正输出。在现实生活中若消费者拥有

良好交易记录,平台就会与消费者保持一致立场,以快速化解纠纷为主要目的,减损公平。其次,数字平台不同纠纷治理方式的关系错置。原本数字平台为解决不同种类网络违法行为造成的法律后果提供多种纠纷化解机制,但是如今平台裁决与大众评审的选用出现关系混同。多数平台并未根据不同治理方式的差异化定位进行案件分类,造成纠纷选用标准混同。最后,法律对于数字平台预防性义务的履行边界规定不清。在上述“人脸识别案”中,法律对于数字平台预防义务的履行标准模糊,同样是平台治理行为滞后的重要原因。例如相关法条并未言明数字平台应承担的“相应责任”究竟是综合责任还是单一责任,学界和司法实践中亦存在不同看法[16]。综上,可以发现对于平台违反该种义务应承担何种具体责任,法律和数字平台均并未明确规定后果,这也为平台预防性义务的履行带来风险。

2) 忽视平台监管义务的履行

鉴于数字平台的空间性,其是面对网络违法行为的第一主体。互联网平台在大数据经济下具有信息、技术等特定的优势与便利,因此其天然的具有监管义务。

通常情况下,平台作为市场组织者,为商家提供网络平台服务,由于平台自身不是网络交易当事人,因此其地位相对中立,趋于维护公共利益[17]。但在特定情况下,平台的监管责任与其私主体身份可能存在矛盾。首先,在与自身利益无关的治理中,平台有可能怠于履行监管职责。如果网络违法行为涉及侵害第三人利益,平台基于维护平台信誉和用户满意度方面的考虑,忽视该行为的发生。其次,平台还同时具有经营者的身份,为了满足对利润的追求,可能忽视监管义务履行。在前述“北京爱奇艺科技有限公司与吴某网络服务合同纠纷案”中,爱奇艺公司既是经营主体又是监管者,在特定情况下,公司为保障自身盈利目标实现可能忽视对于消费者的权益保护并侵害其合法权益。最后,平台在监管过程中具有较大裁量权。一方面,基于网络主体行为的复杂性,法律法规难以对数字平台的监管义务进行列举规定,更多情况下平台需要自行把握义务履行的限度,既达到规制违法行为的目的又避免触犯到法律规定的红线,这在客观上为数字平台带来困难。另一方面,各个平台主体对于法律规范的理解性也不尽相同,在相同情况下有很大概率会做出不同的判断,甚至忽视监管义务的履行。

3. 网络违法行为数字平台治理的新路径

在互联网技术日新月异的今天,传统治理模式已难以适应当前的治理需求,亟待探索针对网络违法行为的数字平台治理新路径。本章将从技术赋能、平台义务强化以及法律体系重构三个维度,针对前文提出的技术迭代矛盾、数字平台治理责任虚化、法律体系碎片化等问题,探讨网络违法行为数字平台治理的新路径,实现网络空间的法治化、规范化发展。

3.1. 重构网络违法行为法律治理体系

技术赋能和义务的强化离不开健全的法律框架支持。目前,我国的互联网治理法律体系存在碎片化和标准不统一等问题,难以形成协同一致的治理合力。因此,亟须从立法层面进行系统性重构,明确认定标准,规范权力边界,构建科学严密的法治网络。

3.1.1. 明确网络违法行为的司法认定标准

网络违法行为类型多样、演变迅速,现行法律在部分领域的界定尚显模糊,网络违法行为认定标准不一,导致司法实践中出现同案不同判、定性困难等现象。立法机关出台互联网平台软法指导大纲,对互联网平台制定的一系列监管规则做出引导[18]。规则应结合司法实践与技术发展,对各类网络违法行为的构成要件、情节严重程度、量刑标准等进行明确界定。同时,建立动态调整机制,定期修订相关法律法规,及时吸纳新型违法形态,确保法律适用的前瞻性与准确性,还应推动跨部门数据共享与协同办案,统一执法尺度,消除管辖冲突,提升司法效率。

3.1.2. 构建数字平台滥用治理权力的防范机制

在强化平台治理责任的过程中，也需警惕平台借治理之名实行垄断行为侵害用户合法权益。所以，为解决缺少对平台滥用权力的限制性规定这一问题，相关部门主体必须要建立有效的保障措施以规范平台治理权力的行使。平台在采取封号、删帖、限流等措施时，应遵循必要性与适度性要求，避免过度干预用户权利；同时，平台要建立申诉复核机制，保障用户的知情权、陈述权与救济权，确保处理决定公正透明。此外，有关主管部门应加强反垄断和反不正当竞争执法，严厉打击平台经营者利用市场支配地位排除竞争对手或实施“独家交易”等行为。对平台滥用权力侵害用户合法权益的，应承担民事赔偿责任；对违法行为严重、扰乱网络生态秩序的，依法承担行政责任；对构成犯罪的，依法追究刑事责任，从而维护公平、竞争、有序的市场环境。

3.2. 技术赋能违法行为治理

技术是导致传统监管理念滞后的根源。机器的学习技术具备适应性与目标取向，底层代码会不断变化。在这种情况下，传统静态的监管理念将无法正确理解和有效捕获数字平台的技术手段[19]。面对日益复杂的网络违法形态，单纯依靠人工审核、传统监管这样的监管手段显然力不从心，数字平台应当学会充分利用大数据、人工智能、区块链等前沿技术，加强平台对非法活动的监测、及时预警和应对能力。

3.2.1. 完善网络算法设计

Table 1. Algorithm evaluation process flowchart
表 1. 算法评估流程表

评估阶段	核心任务	具体内容
1. 风险识别	明确算法目标与潜在风险	识别算法可能对识别网络非法活动(如网络诈骗)、用户权利(如个人信息隐私)和社会公共利益(如信息茧房、歧视、网络谣言)造成的潜在负面影响。
2. 数据审计	审查训练数据的质量与偏见	检查用于训练算法的数据集是否存在偏差、歧视性内容或不完整性，确保数据来源的合法合规。
3. 模型测试	进行公平性与准确性测试	在模拟环境中测试算法对不同群体、不同类型内容的处理结果，评估其识别违法信息的准确率和误判率。
4. 透明度报告	编制算法逻辑说明文档	以通俗易懂的方式向监管机构和公众解释算法的基本逻辑、决策依据和主要参数，打破“黑箱”。
5. 持续监测	建立上线后动态监测机制	算法上线后，确保相关主体持续监测其运行效果，收集用户反馈，定期复盘并优化算法模型，防止出现新的风险。

网络算法是数字平台运行与治理的核心基础，算法机制的缺陷是导致平台无法有效识别、拦截网络违法行为的重要原因。有必要完善网络算法设计，将其作为强化平台对非法活动监测、预警与应对能力的核心引擎。依靠物联网、互联网等数据资源，对网络犯罪及各类风险进行全面采集、实时感知、科学研判、精准处置，以数据控制提升治理的可识别性[20]。为确保算法设计符合治理要求，平台在部署新的内容审核或推荐算法前，可以先行完成以下评估流程(见表 1)，将算法设计、负责及评估主体分离，让社会公众与公权力同时参与充分的交换风险信息并交流反馈意见，将算法权力的异化扼杀在摇篮中[21]。

3.2.2. 强化平台智能管理功能

除了优化网络算法设计，平台还需在管理功能上加大投入，构建全方位、智能化的管理体系，防止

平台治理行为滞后。要完善平台的智能拦截与处置功能,针对已明确的网络违法行为,设置智能拦截规则,对违法内容进行即时屏蔽、对违法账号进行限权或封禁,实现对违法行为的快速处置;对于疑似违法行为,采取智能限流、弹窗提醒等措施,降低其传播范围。并且注意防止形成数据孤岛,数据孤岛是影响数据挖掘和使用的关键因素,也是在大数据监督工作中最为受关注的一个问题[22]。建立平台间的技术共享与数据互通机制,打破数据孤岛,实现对跨平台网络违法行为的协同治理,让网络违法行为在网络空间中无处遁形。

3.3. 发挥数字平台守门人义务

鉴于数字平台在数字市场中发挥的关键作用,它们应成为市场的守护者,并承担起维护数字市场秩序的更大责任。作为网络服务的守门人,平台在享受流量带来的红利与其自身技术优势的同时,必须履行相应的社会责任与法律义务。这意味着,数字守门人应承担与其权利相匹配的特殊义务,利用自身的数字权威与能力,防范并制止其控制领域内不法或不当行为对用户权益的侵害[23]。

3.3.1. 前置平台审查责任

传统治理模式下,平台往往遵循“通知-删除”规则,即在收到投诉或监管指令后才采取行动,这种滞后性使得许多违法行为在对被害者造成实质损害或引起严重后果后才被制止。这种事后审查的弊端在于因案而异的方式无指导性的普适效力,也无法从源头上解决系统性的问题[24]。

新路径要求将审查责任前置,强化事前监管,以防平台忽视履行预防性义务。为此,建议构建一个基于“内容风险等级”与“发布者信誉等级”的差异化、分层分类审查模型。该模型将平台内容划分为“高风险”“中风险”与“低风险”三个等级,并对发布者进行信誉进行划分:“高风险发布者”、“中风险发布者”和“低风险发布者”。根据不同组合采取差异化的审查策略,实现精准治理。内容风险等级建议平台根据国家相关法律法规和行业规范,预先设定关键词库和内容特征库,通过 AI 进行自动分类。而发布者信誉等级则是根据发布者的实名认证情况、历史发布内容合规记录、用户举报次数、账号活跃度等多维度数据进行动态评分和分级。

3.3.2. 加强针对平台的监督审查机制

赋予平台审查责任的同时,必须建立强有力的外部监督机制,防止平台不履行或滥用监管权力。监管机构的主要监督职责应明确界定,相关部门应建立协调监督的机制。定期对数字平台的监管治理工作进行检查评估,重点审查平台的监管审查制度建设、违法内容处置、风险防范措施等落实情况,对未履行监管审查责任、治理不力的平台,依法采取责令整改、罚款、责令停业整顿等行政处罚措施,并及时将处罚结果向社会公示,形成震慑。同时,引入社会监督与公众参与机制,搭建可以治理投诉举报的平台,畅通公众对平台内网络违法行为以及平台不作为行为的投诉举报渠道。通过建立一个鼓励民间社会积极参与的社会治理平台,旨在引导公众参与平台监督与治理的各个层面和环节,从而营造出互利共赢、协同创造、共享成果的文化氛围和社会治理格局。

致 谢

在此对项目资助者、对文章提供指导和帮助者、给予转载和引用权的文献、研究思想和设想的所有者,表示感谢。

基金项目

本文由浙江农林大学大学生科研训练项目资助,属于 2026 年省级大学生创新训练项目,课题编号 S202510341124。

参考文献

- [1] 周濂. 现代政治的正当性基础[M]. 上海: 上海三联书店, 2021.
- [2] 马怀德, 王晓淑. 数字平台参与政府治理的正当性及其实现路径[J]. 治理研究, 2024, 40(6): 25-39, 157-158, 2.
- [3] 陈荣昌. 网络平台数据治理的正当性、困境及路径[J]. 宁夏社会科学, 2021(1): 72-80.
- [4] 陈星. 论网络空间主权的理论基础与中国方案[J]. 甘肃社会科学, 2022(3): 113-121.
- [5] 郝英兵. 刑罚正当性根据研究[M]. 北京: 法律出版社, 2019.
- [6] 周尚君, 罗有成. 数字正义论: 理论内涵与实践机制[J]. 社会科学, 2022(6): 166-177.
- [7] Rudin, C. (2019) Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, 1, 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- [8] 冷传莉, 王宗峰. 平台经营者管理权的滥用与规范进路[J]. 南昌大学学报(人文社会科学版), 2024, 55(4): 93-104.
- [9] 唐俊麟. 平台处罚仅是“依合同的自治”?——法律关系视阈下平台惩戒性管理行为的效力探析[J]. 交大法学, 2026(1): 133-147.
- [10] 刘河庆, 梁玉成. 透视算法黑箱-数字平台的算法规制与信息推送异质性[J]. 社会学研究, 2023(2): 49-71.
- [11] 印波, 庄新宇. 短视频算法风险治理规范的运行难点与完善路径[J]. 中国特色社会主义研究, 2025(3): 81-94.
- [12] 程雪军, 侯姝琦. 互联网平台数据垄断的规制困境与治理机制[J]. 电子政务, 2023(3): 2-18.
- [13] 耿宇. 算法黑箱法律规制路径的比较分析[D]: [硕士学位论文]. 苏州: 苏州大学, 2024.
- [14] 李志刚, 李瑞. 共享型互联网平台的治理框架与完善路径——基于协同创新理论视角[J]. 学习与实践, 2024(4): 76-83.
- [15] 李朝. 我国预防性平台纠纷治理的实践逻辑与优化路径[J]. 法商研究, 2025, 42(4): 68-84.
- [16] 倪楠, 丁元. 电商平台经营者违反安全保障义务的“相应责任”解构——以比例连带责任为中心[J]. 上海财经大学学报, 2025, 27(4): 138-152.
- [17] 金美蓉, 李倩. 错误成本分析理论下互联网平台准监管责任的问题与完善[J]. 中国矿业大学学报(社会科学版), 2023, 25(2): 96-116.
- [18] 叶明, 贾海玲. 双重身份下互联网平台自我监管的困境及对策——从互联网平台封禁事件切入[J]. 电子政务, 2021(5): 12-20.
- [19] 孙晋. 数字技术规范法律化与数字平台善治[J]. 中外法学, 2025, 37(5): 1283-1302.
- [20] 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.
- [21] 赵一丁, 陈亮. 算法权力异化及法律规制[J]. 云南社会科学, 2021(5): 123-132.
- [22] 宋伟, 邵景均. 基于大数据的廉政监督模式与发展趋势分析[J]. 中国行政管理, 2019(7): 26-30.
- [23] 解正山. 约束数字守门人: 超大型数字平台加重义务研究[J]. 比较法研究, 2023(4): 166-184.
- [24] 李本, 徐欢颜. 中国方案: 欧盟对数字平台“守门人”监管革新的启示[J]. 上海大学学报(社会科学版), 2023, 40(5): 105-119.