

短视频沉迷中的智能推荐算法反思与规制

杨美玲, 潘 静

(北京邮电大学 人文学院, 北京 100876)

摘要: 短视频沉迷现象的蔓延与智能推荐算法的深度应用密切相关, 已成为影响个体发展与社会治理的重要议题。智能推荐算法通过“数据画像—强化推送—即时反馈”的技术闭环持续捕获用户注意力, 由此诱发系统性公共健康风险。现行治理框架在算法层面存在显著滞后性, 突出表现为算法责任认定标准模糊、立法供给不足以及监管机制的碎片化。因此, 短视频沉迷中智能推荐算法的治理应以技术特性为基础、社会影响为导向、法律回应为保障的阶梯式治理路径: 在法律层面, 明确沉迷认定标准, 构建算法风险分级与三元责任体系; 在监管层面, 建立算法动态监测机制, 推行监管沙盒与阶梯式干预; 在多元共治层面, 完善行业自律, 畅通公众参与渠道, 推进国际规则对接。希冀为破解算法驱动型沉迷提供实操路径, 也为数字时代技术风险的法律规制提供范式参考。

关键词: 短视频; 沉迷; 智能推荐算法; 法律规制; 多元共治

中图分类号: D922.16

文献标识码: A

DOI: 10.19358/j.issn.2097-1788.2025.11.013

引用格式: 杨美玲, 潘静. 短视频沉迷中的智能推荐算法反思与规制 [J]. 网络安全与数据治理, 2025, 44(11): 79-85, 92.

Reflection and regulation of intelligent recommendation algorithms in short video addiction

Yang Meiling, Pan Jing

(College of Humanities, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: The proliferation of short video addiction is closely associated with the deep-seated application of intelligent recommendation algorithms, emerging as a significant issue affecting individual development and social governance. Intelligent recommendation algorithms continuously capture users' attention through the technical closed-loop of "data profiling—enhanced push—instant feedback", thereby inducing systemic public health risks. The current governance framework shows remarkable lag at the algorithmic level, mainly manifested in vague criteria for algorithmic liability determination, insufficient legislation, and fragmented regulatory mechanisms. Therefore, the governance of intelligent recommendation algorithms in short video addiction should adopt a ladder governance path based on technical characteristics, guided by social impacts, and guaranteed by legal responses: at the legal level, clarify the standards for identifying addiction, and establish the algorithm risk classification and three-party liability system; at the regulatory level, establish a dynamic monitoring mechanism for algorithms, and implement regulatory sandboxes and stepwise interventions; at the multi-stakeholder co-governance level, improve industry self-regulation, smooth channels for public participation, and promote the alignment of international rules. It is expected to provide practical approaches for tackling algorithm-driven addiction and offer paradigmatic reference for the legal regulation of technological risks in the digital era.

Key words: short video; addiction; intelligent recommendation algorithm; legal regulation; multi-stakeholder governance

0 引言

短视频是以连续画面、背景音乐及字母等组成的区别于传统长视频（影视剧、综艺等）的内容载体^[1]。中国互联网络信息中心（CNNIC）的统计数据显示，截至

2024年12月，我国短视频用户规模已达10.40亿，网民使用率高达93.8%^[2]，位居各类互联网应用前列。《中国网络视听发展研究报告（2024）》数据亦显示，人均单日短视频使用时长已达151 min，远超其他互联网应用。

过度使用短视频不仅影响学业和工作,更可能造成认知失调和情绪波动,甚至存在间歇性停工的潜在危害^[3]。智能推荐算法作为短视频平台实现用户增长的核心技术,其对沉迷现象的驱动作用尚未得到充分规制。我国既有研究多聚焦于算法责任认定、平台义务界定、未成年人保护措施以及算法监管立法完善等方面,主张通过加强算法透明度、强化分级管理等手段规制平台行为。然而,这些结论往往存在将算法技术“工具化”而忽视其沉迷诱导逻辑,容易简化平台责任认定,同时现有法律体系下结构性缺陷分析不足,导致对策建议的实操性有限。

据此,有必要围绕智能推荐算法与短视频沉迷的交互关系展开研究。首先,解构智能推荐算法的技术特性,阐明其通过神经心理学机制强化用户依赖的本质,论证法律规制的必要性。其次,剖析当前治理困境,包括沉迷判定标准缺乏、“立法+行政”双轨制执行效果不彰、监管与自律协同失效等,揭示传统规制手段适配短板。最后,从法律规范完善、监管模式转型和多元共治机制三个层面,提出针对性治理方案。这不仅为破解算法驱动型短视频沉迷提供实操路径,更为数字时代技术风险的法律规制提供范式参考,推动从“被动应对”到“主动构建”的治理转型。

1 智能推荐算法的作用机制及伦理规制需求

1.1 算法助推机制:沉迷行为的技术强化路径

智能推荐算法作为信息过滤程序,通过分析系统规则、用户信息及人机交互痕迹,构建用户画像,依靠内容标签向特定用户精准分发更大概率被关注的项目(产品或服务)(见图1)。目前短视频平台广泛应用的协同过滤、混合推荐等算法模型,虽本身为中性技术工具,并无价值预设,但在商业公司主导的算法模型中,利润至上的价值取向使其成为一种内嵌商业逻辑的“伪中立”技术^[4]。平台为追求用户规模扩张和使用时长增加,将算法作为实现该目标的关键手段。算法推荐模型的核心优势在于能够“永久计算”和“永久算计”^[5],用户与算法之间的交互构成一种递归关系^[6]:用户行为数据持续反哺算法迭代,而算法通过深度学习不断提高推送内容的“定制化”程度,诱导用户沉溺于所构建的“兴趣泡沫”之中。长此以往,个人真正的自由将受到威胁^[7]。

从行为心理学视角看,这种技术机制对用户认知产生显著干预。用户在观看个性化短视频时,大脑默认模式网络(Default Mode Network, DMN)激活,且与视觉、听觉通路的耦合增强,同时削弱楔前叶和扣带皮层的调节功能,导致注意力资源过度集中于短视频内容,注意力调节能力持续下降,认知控制功能受到抑制,最终可

能导致短视频使用行为失控^[8],此即算法助推下沉迷行为的核心形成机制。

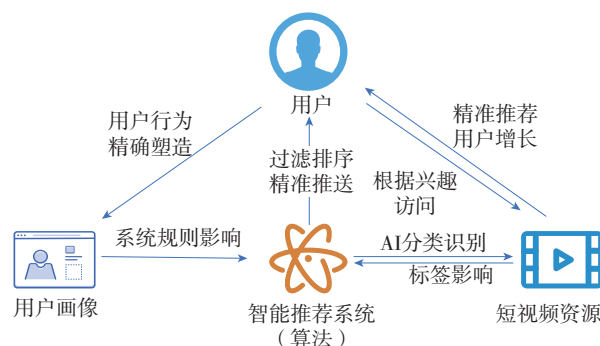


图1 智能推荐系统核心

1.2 算法伦理的失范表现与规制需求

从算法伦理角度而言,国际社会虽通过诸如《美国人工智能研究和发展战略规划》(The National Artificial Intelligence R&D Strategic Plan)、联合国《机器人伦理初步报告草案》、欧洲政治传播中心《人工智能时代:确立以人为本的欧洲人工智能战略》(The Age of Artificial Intelligence: Towards a European Strategy for Human-Centric Machines)、欧盟《可信人工智能伦理指南》《人工智能法》(Artificial Intelligence Act)等^[9]文件确立算法应遵循的“以人为本”的原则,但软法属性使其难以约束平台的逐利倾向。短视频算法设计中,“用户时长最大化”目标凌驾于健康使用原则之上,如通过“无限滚动”“即时反馈”等功能刻意延长用户停留时间,构成对自主选择权的隐性剥夺。这种伦理失范亟需刚性规范介入,确立算法设计的公共利益优先条款。

设计伦理学家特里斯坦·哈里斯(Tristan Harris)指出,用户沉迷并非源于意志力薄弱,而是因为屏幕后面有数千人努力工作,为的就是破坏你的自律^[10]。平台技术开发者理应遵循行业伦理准则,避免设计诱导用户过度使用或成瘾为目的的功能,并在算法运行效果的持续监测与评估中,及时采取干预措施。但算法的“黑箱特性”掩盖了其内嵌的价值判断,平台通过模糊代码逻辑、隐藏数据处理流程,既规避公众监督,又将开发者责任稀释于技术抽象化过程中。对此需建立算法透明度规制,要求平台披露推荐逻辑中与沉迷相关的参数设置,如推送频率、内容刺激性阈值等。

智能推荐算法的设计亦应充分考量人性化原则,即技术设计不仅关注算法效率与效果,更需尊重用户的个体差异与心理需求。算法模型对用户短期兴趣的过度捕捉,导致内容推送呈现“同质化锁定”特征,忽视长期

心理健康需求。心理学研究表明,持续接收单一类型刺激会削弱用户的认知多样性与社会适应能力。这要求规制框架应增设“算法向善”义务,强制平台在推荐模型中嵌入健康使用指标,如内容多样性权重、使用时长预警参数等。算法的核心目标不仅在于提升用户粘性,更应保障用户的自主性与长远福祉,促进技术与人文的和谐共生。

此外,智能推荐算法在实现平台商业价值的同时,也面临平衡社会价值的难题。算法对情感化、娱乐化内容的偏好推送,加剧了信息茧房与认知极化,削弱公共话语空间的理性基础。平台作为社会信息枢纽,未履行内容生态治理义务,反而利用算法放大内容的传播效应。对此应通过规制明确平台的社会治理责任,要求其建立算法推荐的公共价值评估机制。通过技术手段与伦理考量相结合,确保推荐算法在实现商业利益的同时,不侵蚀公共价值和社会伦理,实现二者的有机统一,为短视频沉迷治理提供价值基础与规范依据。

2 短视频沉迷治理的规范缺陷与治理困境

2.1 规范要件解构:沉迷行为的法律界定与认定标准缺失

短视频沉迷治理的首要障碍在于法律规范对基础概念的界定模糊。学界对“沉迷”与“成瘾”的语义纠缠尚未厘清:前者侧重社会行为层面的过度依赖,后者带有医学病理标签,而现行法律多将二者混同。美国心理学家 Ivan Goldberg 基于《美国精神疾病分类与诊断手册》DSM-IV 中关于药物成瘾的判断标准提出互联网成瘾障碍(Internet Addiction Disorder, IAD)的定义^[11]。但“成瘾”一词带有病理性标签,与酒精或药物依赖相关联,需对患者进行临床治疗。因此,有学者认为不应简单地将二者等同,社交媒体过度使用所造成的沉迷问题医学化尚有榷论^[12]。还有一些学者的研究中,虽然关注了短视频沉迷问题,但仍未给出相关定义^[13]。中国国家卫生健康委员会发布的《中国青少年健康教育核心信息及释义(2018版)》将网络成瘾定义为“无成瘾物质作用下对互联网使用冲动的失控行为”^[14],但仍局限于疾病范畴,与短视频沉迷的社会行为属性存在本质差异。在综合以上观点及政府相关规范文件用语后,鉴于“成瘾”一词可能所蕴含的潜在负面含义,笔者选择“短视频沉迷”这一更为中性的词语进行论述。将其理解为个体过度依赖和沉溺于短视频内容,严重影响学习、工作、生活等现实社会活动^[15],并损害身心健康的行为现象^[16]。

目前,短视频沉迷判断标准的探讨主要集中于以下两个方面:其一,个体每日观看短视频的时间远超正常娱乐需求,无法观看时感到不安或失落,产生焦虑、烦

躁等情绪反应,对其产生心理依赖,并伴随强迫使用观念和自我控制能力减弱等^[17];其二,个体因网络虚拟空间与现实生活脱轨,继而忽视日常责任和义务,如学业、工作、家庭等^[18],造成人际关系紧张,社交能力下降,甚至因沉迷短视频而引发社会性功能受损。然而,现有标准在具体应用中存在一定的局限性。首先,“每日观看时长”的量化标准因个体差异丧失普适性,“正常娱乐需求”的主观判断也会导致执法尺度混乱。其次,“社会性功能损伤”的界定缺乏操作指引,学业受影响、人际关系紧张等后果与短视频使用的因果关联难以举证。这种规范要件的模糊性,也使得如《中华人民共和国未成年人保护法》第七十四条“不得提供诱导沉迷的服务”等条款沦为宣示性规定,无法形成有效的法律评价体系。

2.2 制度供给检视:现行法律规制的结构性缺陷

从比较法视角观察,各国已形成差异化的规制路径:欧盟采取严格的事前监管模式,美国侧重行业自律与司法救济并重,我国则采取“立法+行政”的双轨制模式。尽管各国政府和社会各界已认识到此问题的严重性,并尝试通过立法和监管手段加以解决,但现有法律制度在应对此问题时仍然存在诸多不足。

2.2.1 规范内容的滞后性

法律滞后性是法律与现实之间的时空差异性及由此引起的法律功能性损伤^[19]。法律规范对短视频沉迷的回应始终落后于技术迭代速度,各国立法仍延续对网络游戏的规制路径。我国自2001年至2017年,针对未成年人网络保护制定颁布了共47项相关法律法规和政策,包括综合政策和针对网吧管理、游戏时间限制、网络游戏实名制、网络游戏分级、家长监护工程等专项政策^[20]。2024年11月,国家网信办发布《移动互联网未成年人模式建设指南》,涉及短视频平台的内容只是着重在风险治理、内容建设等维度对未成年用户加以保护,对沉迷问题仅提及“不得提供诱导其沉迷的产品和服务,及时修改可能造成沉迷的内容、功能和规则。”^[21]韩国制定了“防治中小学生网络游戏成瘾第一个三年总体规划”(The Master Plan to Prevent and Solve Internet Addiction)和“第二个防治网络成瘾的总体规划”(The Second Master Plan to Prevent and Solve Internet Addiction)^[22]。日本政府也颁布了《加强青少年网络环境安全法》《交友类网站限制法》等法案。总体而言,各国涉及互联网使用的保护人群及保护范围主要集中于网络游戏领域,将短视频沉迷简单纳入“网络成瘾”框架,忽视其算法驱动的特殊性。即使部分国家开始关注智能推荐算法对沉迷行为的助推作用,如欧盟委员会公布的《数字服务法案》(Digital Services Act, 2022)要求大型在线平台必须提供“透明算

法”，但未将沉迷风险纳入评价体系，凸显规范内容与治理需求的错位。

2.2.2 实施机制的体系性失衡

短视频沉迷问题涉及多个主体，包括公权力机构的调控作用、短视频平台应当承担的义务、内容创作者应遵守的内容上传规则以及用户在何种情况下需自我承担风险等。然而，现行法律对这些主体之间的责任分配体系显著失衡。大多数国家的互联网法律体系以平台责任为核心。例如，我国2020年施行的《网络信息内容生态治理规定》第八条规定“网络信息内容服务平台应当履行信息内容管理主体责任”，2021年国家互联网信息办公室发布的《关于进一步压实网站平台信息内容管理主体责任的意见》对网站平台信息内容管理主体责任进行了详细规定，同年国家市场监督管理总局发布了《互联网平台落实主体责任指南（征求意见稿）》，从34个方面拟全面规定平台的主体责任，要求平台积极发挥主观能动性，合理约束自身需求。但每个主体责任边界未通过法律明确，出现“谁都该负责却谁都不担责”的局面。算法设计中“诱导沉迷”的主观意图也难以举证，过度收集数据与个性化推送的合法性边界无法区分。同时用户沉迷是算法推送、个体心理与社会环境共同作用的结果，技术介入的比例难以量化，因果关系证明困难，传统侵权责任框架难以适用。

各国目前应对短视频沉迷的手段主要集中在技术防护和行业自律方面。在执行能力上，算法的“黑箱特性”构成难以逾越的技术壁垒，代码迭代速度远超备案更新频率，模拟测评环境与真实用户场景存在本质差异，导致监管部门无法有效监测推荐逻辑。各监管机构在管理体系中出现失位问题，相关规定形同虚设。早在2007年，我国已发布《关于保护未成年人身心健康实施网络游戏防沉迷系统的通知》，并要求各网游平台和视频平台执行，但该举措并未得到有效落实。此外，国家互联网信息办公室、工业和信息化部、公安部、国家市场监督管理总局四部门于2022年1月联合发布《互联网信息服务算法推荐管理规定》，文件指出监管主体为双层次、多部门联合监管，但监管机构既包括国家网信部门、公安机关、市场监管，又包括各级人民政府及其主管部门等，各机构的法律地位各不相同，监管权限、对象及内容也无明确规定。这种技术壁垒叠加“运动式执法”的短期性，使得《移动互联网未成年人模式建设指南》中的防沉迷要求沦为“纸面上的义务”，54.1%的未成年短视频用户长期暴露在无有效监管的使用环境中^[23]。

2.2.3 国际协同的规制断层

在防治短视频沉迷问题上，各国规制模式呈现明显

分化，亚洲国家采用“立法+行政”的强干预模式，但不足之处在于政策以及措施的推广性与实践性较差，难以取得预期效果。欧美各国则主要通过向非政府组织和社会非营利性机构提供资金的方式，开展咨询、服务、治疗等大规模、全员式的项目活动来推动防治工作的实践。但由于举办大型活动投入成本过高，可持续性有待进一步考量。国际性文件等也仅确立原则性共识，未形成具有约束力的跨境监管机制，导致平台可通过注册地选择规避严格管理，形成“监管套利”空间。这种国际协同的缺失，与短视频算法的全球化部署形成鲜明对比，加剧了治理碎片化。各个国家应结合实际情况，借鉴他国经验以补己之短，实现“互联网+”的无限可能。

综上，掣肘法律制度在治理短视频沉迷问题的原因是多方面的，单纯依靠解决某一问题难以有效推进治理。因此需探寻智能推荐算法在短视频沉迷问题中法律层面的深层因素，通过制度设计解决该问题，以便更有效地应对这一日益严重的社会问题。

2.3 治理效能反思：多维矛盾的现实呈现

2.3.1 技术逻辑与规制能力的冲突

智能推荐算法的“不可解释性”与规制所需的“透明度”形成根本矛盾。算法通过深度学习形成的非线性决策路径，使得即使专业人士也难以完全解码其推荐逻辑。平台虽按要求披露算法原理，但代码的快速迭代与商业秘密保护诉求，导致公开内容多为“脱敏后的技术描述”，无法反映与沉迷相关的核心参数。在传统行业监管中，相关部门可以直接进入经营场所查看实际流程，但是在平台应用推荐算法过程中，监管部门无法通过公开透明的渠道进行监测。这种信息不对称使得监管部门陷入“既无法监测算法运行，又难以评估干预效果”的困境。智能推荐算法的技术逻辑及应用方式所带来的不可解释性，导致算法“黑箱”难以被公众理解，对大众的知情权及自主决策权构成挑战，进而产生算法不可监督、难以追责等治理困境^[24]。

2.3.2 商业利益与公共责任的失衡

算法治理在很大程度上依赖于技术规程和操作控制，因此基于主体责任的算法自律机制至关重要^[25]。行业自律强调平台在运营过程中自觉遵守行为规范，提升内容质量与用户体验。但平台的逐利本性与防沉迷义务存在结构性冲突。智能推荐算法在一定意义上具备互联网信息资源配置的公共权力，并具有一定的意识形态属性^[26]。比如，短视频行业的“注意力经济”模式将用户时长直接转化为广告收益，导致算法设计天然倾向“沉迷强化”。“无限滚动”功能利用间歇强化原理延长使用时间，

“点赞反馈”机制通过多巴胺刺激塑造使用习惯。这种商业逻辑使得行业自律沦为象征性合规，即使平台设置时长提醒，也多采用弱化设计降低用户感知。当利润最大化成为算法的核心目标时，社会责任只能成为“装饰性条款”。

2.3.3 主体认知与治理需求的落差

公众对算法风险的认知不足，以及对自身行为后果的判断失误，是短视频沉迷治理中的重要问题。现代法律高度重视个人意思自治，其发挥个人主观能动性的重要前提在于人“有能力在给定的各种可能性的范围内，自主地和负责地决定他的存在和关系、为自己设定目标并对自己的行为加以限制”^[27]。但对公众而言，“个性化推荐—行为数据收集—算法迭代”的闭环将用户偏好转化为技术规训工具，导致用户陷入“技术无意识”状态。即使平台提供算法关闭选项，信息茧房也将个体禁锢在同质化的回音室之中，使得思想市场沦为算法导演的假象^[28]。用户也因缺乏对危害的认知而放弃使用，使得《中华人民共和国个人信息保护法》所赋予的知情权与选择权难以实现。用户在此情形下往往难以完全意识到个人数据的具体使用方式与范围，特别是算法对个人数据的最终使用情况可能“已经远远超出个人的意图范围和认知能力”^[29]。如此一来，过度收集个人数据甚至数据滥用的现象不仅引发公众反感，还可能侵犯个人隐私。

在我国算法治理共同体中，监管机构、企业、社会组织、公众都是治理主体，都是有其责亦应尽其责的“一元”，都是多元共治格局的一个“方阵”。其中，监管居于“中枢”地位^[30]。但“监管机构—平台—用户”的这种流线型监管路线存在平台责任不清的风险^[31]。加之算法专业知识的匮乏导致规制措施“水土不服”，如要求平台披露算法代码却忽视代码迭代的动态性，这种主体认知与治理需求的落差，使得治理体系缺乏有效的社会基础与实施能力。

3 短视频沉迷的治理路径优化：基于问题导向的规制调整

智能推荐算法引发的短视频沉迷问题，本质上反映了技术理性与人文价值之间的深刻矛盾。算法已经不仅仅是技术工具，更成为一种新型的社会权力形态。应推动建立“预防—控制—救济”三位一体的系统性规制框架，通过立法完善、监管创新和社会共治的协同配合，实现技术发展与社会福祉的有机统一。

3.1 法律规范体系的完善

3.1.1 沉迷行为的法律认定标准具象化

针对短视频沉迷缺乏统一认定标准的问题，可以在

法律条文修订中确立“双维度认定框架”。客观上采用“时长+频率”的复合标准，如连续使用时长超过45 min且日均累计超120 min（未成年人60 min）。主观上引入“功能性损伤”量化指标，包括学业成绩下滑幅度（如学期成绩下降25%以上）、社交活动参与频次减少（如线下社交每周不足3次）等可举证事实，明确“兴趣泡沫”引发的认知窄化属于损害情形。

3.1.2 算法设计的义务性规范嵌入

算法伦理失范中存在“商业利益优先”问题，应根据智能推荐算法的“注意力捕获强度”建立三级风险评级体系：一级（低风险）适用于日均用户时长<30 min的平台，仅需履行基本告知义务；二级（中风险）适用于日均用户时长为30~90 min的平台，需增设使用时长可视化功能；三级（高风险）适用于日均用户时长>90 min的平台，要求强制嵌入“三重机制”——使用时长预警、内容多样性配置、未成年人模式强制开启选项。

该分级制度需与《互联网信息服务算法推荐管理规定》第十二条衔接，明确高风险算法的强制备案义务，由国家网信部门组织的算法评审委员会进行合规性审查。这种基于技术机理的分级规制，能够有效破解算法黑箱导致的规制被动性。同时参照欧盟《人工智能法》第三章对“高风险人工智能系统”的分类标准，将未履行上述义务的算法界定为“高风险算法”，面临相关法律法规规定的行政处罚。

3.1.3 三元主体的责任边界厘清

为破解责任分配混乱，立法应明确“算法设计者—平台运营者—内容提供者”的三元责任架构，确立“权责利”对等的责任体系。算法设计者对“诱导性参数设置”承担过错推定责任，若无法证明参数设置的合理性，需承担30%~50%的赔偿责任。平台运营者对算法运行的实时监测承担无过错责任，即使已履行形式提醒义务，若未通过技术手段阻断超时长使用，仍需承担主要赔偿责任。内容提供者对推送内容的“成瘾性特征”，可适用《网络信息内容生态治理规定》第三十四条规定，采取限制性处置措施。通过技术环节与责任节点的精准绑定，解决“多因一果”场景下的责任划分难题。

现行法律体系对算法规制的局限性在于过度关注行为结果而忽视技术特性。人工智能兼具多种技术特性，直接移植适配域外经验并不可取，需要建立符合中国数字社会发展阶段的法律框架。

3.2 监管机制的技术性转型

3.2.1 算法透明化监管的技术实现

人工智能法律治理的有效性取决于能否适配人工智

能的技术特性^[32]。针对算法黑箱导致的监管被动,可以由国家网信办牵头建立“算法备案与动态监测系统”。短视频平台需在上线前提交推荐算法的核心参数图谱,系统通过标准化API接口实时抓取用户行为数据流(脱敏处理后),运用机器学习模型自动识别“沉迷诱导特征”。检测结果直接关联行政处罚。对于不合规平台,将其列入“算法失信名单”,限制其参与政府采购与评优评先,破解“运动式执法”的短期性缺陷。

3.2.2 推行“监管沙盒”试点制度

“监管沙盒”机制为算法科技创新提供了一个相对宽松的“安全空间”,允许企业在一定时间和范围内测试新技术,同时确保消费者保护和市场稳定,是解决商业利益与公共责任冲突的有效尝试。可参照北京、上海等地已有的算法沙盒经验,在短视频用户密集地区开展试点,对平台新型防沉迷技术,允许限定用户规模内测试6个月,测试期内可豁免部分合规要求,但需向监管部门实时传输沉迷率变化数据。对试点中证明有效的技术方案,形成“创新—验证—推广”的良性循环。探索构建符合我国国情、与国际接轨的算法科技创新监管工具暨中国版“监管沙盒”。

3.2.3 建立阶梯式干预机制

需要注意的是,监管转型应当遵循比例原则,建立“阶梯式”监管机制。对轻度违规(如提醒功能隐蔽)采取约谈整改;对中度违规(如未成年人保护模块失效)实施罚款并公开曝光;对重度违规(如刻意规避监管检测)暂停算法推荐功能。并且对创新性防沉迷设计给予政策激励,形成“刚性底线”与“柔性引导”相结合的监管格局。如此既解决了监管弹性不足问题,又通过梯度处罚强化威慑力。

法律监督在信息化、智能化和大数据的基础上,已经迭代升级步入数字化阶段,固守传统法律监督方式显然不能适应数字时代要求,必须主动掌握和运用数字技术,进行法律监督数字化变革^[33],构建一个“技术驱动、数据赋能、协同高效”的新型监管体系。这是“适应数字时代发展新趋势,实现监督高质量发展,助力法治监督体系建设效能提升的重要举措”^[34]。

3.3 多元协同的治理生态构建

算法治理的特殊性在于,传统的政府—市场二元结构已经难以应对技术带来的复杂性。数据权益的实现需要平台、用户等多方协同,算法伦理的约束离不开企业、社会的积极参与。因此,必须突破以往的框架,完善政府、企业、社会多元主体分工合作机制,探索多元主体协同治理的制度路径,共建共享数字生态的治理格局^[35]。

3.3.1 行业自律的刚性化改造

中国网络视听节目服务协会发布的《网络短视频内容审核标准细则》代表了行业自律的初步尝试,但具体尚未触及算法设计层面。针对行业自律形同虚设问题,应推动建立由技术专家、法律学者、社会公众代表共同组成的“算法伦理委员会”,对推荐算法进行伦理评估,设置“黑名单”制度,对诱导性设计平台实施行业联合惩戒。同时设立“防沉迷创新基金”,对沉迷率每年下降的平台给予税收减免,通过利益激励破解平台逐利性与社会责任冲突。

3.3.2 公众参与的制度化渠道

为提升用户算法素养,需进一步加强“知情—参与—救济”全链条机制。增设“算法知情权”,用户有权要求平台通过可视化界面说明其推荐逻辑中影响使用时的核心参数。将“技术民主化”理念引入算法治理,建立“算法改进众包平台”,用户可提交防沉迷建议(如夜间模式自动降低推送频率),被采纳建议的提供者可获得平台会员权益(如免除广告服务)。同时,完善公益诉讼制度,扩大公益诉讼主体范围,允许检察机关、未成年人保护组织以“公共利益受损”为由对算法沉迷提起民事公益诉讼。特别需要注意的是还需培养公众算法认知能力,将算法素养教育纳入国民教育体系。通过向公众普及短视频沉迷的内部机制,有望提高公众对该潜在风险的认识和理解,降低短视频沉迷风险以实现“治本”之效果。

3.3.3 国际协同的规则对接

面对跨境治理碎片化,可以依托金砖国家数字经济合作机制,推进治理多边协议的达成,确立未成年人保护最低标准、算法透明度互认准则、沉迷数据跨境共享机制三项共识。也可要求向我国境内提供服务的境外平台,在境内设立算法合规机构,否则限制其在华业务开展。中国应积极参与相关国际规则的制定,建立算法治理的国际对话机制,推动建立公平合理的全球算法治理秩序。

上述三个维度并非相互割裂,而是构成一个有机整体。法律规范体系完善为监管和社会共治提供制度基础,监管创新确保法律规范的有效实施,社会共治则弥补政府监管的不足。三者相互支撑、相互促进,共同构成智能推荐算法的系统性规制框架。

4 结论

智能推荐算法诱发的短视频沉迷现象,并非技术演进中的偶然插曲,而是数字文明转型期必须直面的结构性命题。当平台—算法—资本合谋塑造的“制度性成瘾”

重塑人们的注意力分配模式，这一现象已超越单纯的个体行为偏差，演变为现代性困境在数字领域的典型呈现。

从法哲学视角审视，智能推荐算法的治理必须超越传统规制范式，破除“技术中立”的迷思，正视算法作为新型社会权力的本质。构建“预防—控制—救济”的全过程治理体系，将技术伦理嵌入算法全生命周期。重申“科技向善”的文明底线，使技术发展回归服务人类福祉的本源。放眼全球，算法治理的竞争已超越技术标准，直指制度文明。中国若欲在这一竞争中贡献“法秩序增量”，必须走出“监管—放松”循环，迈向“法律—行政—社会”多层治理，以法律价值为锚、以行政监管为刃、以多元共治为盾。让算法在每一次推送中都必须回答一个古老而常新的问题：技术是否仍在为人而存在？只有当法律能够为这一问题提供制度化的、可强制执行的答案，数字文明方不至沦为“代码利维坦”的新神话。

参考文献

- [1] 12426 版权监测中心. 2021 年中国短视频版权保护白皮书 [EB/OL]. (2021-05-20) https://epaper.chinaxweb.com/epaper/2021-05/20/content_99797650.html.
- [2] 中国互联网络信息中心. 第 55 次《中国互联网络发展状况统计报告》[EB/OL]. (2025-01-17). <https://www3.cnnic.cn/n4/2025/0117/c88-11229.html>.
- [3] Feng Yafei, Li Lifu, Zhao Anqi. A cognitive-emotional model from mobile short-form video addiction to intermittent discontinuance: the moderating role of neutralization [J]. *International Journal of Human-Computer Interaction*, 2022, 40 (6): 1-8.
- [4] 李林容. 网络智能推荐算法的“伪中立性”解析 [J]. 现代传播 (中国传媒大学学报), 2018, 40 (8): 82-86.
- [5] 李龙飞, 张国良. 算法时代“信息茧房”效应生成机理与治理路径——基于信息生态理论视角 [J]. 电子政务, 2022 (9): 51-62.
- [6] GILLESPIE T. The relevance of algorithms [M]. MIT Press, 2013.
- [7] 丁晓东. 论算法的法律规制 [J]. 中国社会科学, 2020 (12): 138-159, 203.
- [8] SU C, ZHOU H, GONG L, et al. Viewing personalized video clips recommended by TikTok activates default mode network and ventral tegmental area [J]. *NeuroImage*, 2021, 237: 118-136.
- [9] 臧志彭, 解学芳. 人工智能时代文化产业主流价值传播: 重塑与建构 [J]. 毛泽东邓小平理论研究, 2019 (4): 48-54, 108-109.
- [10] [美] 亚当·奥尔特. 欲罢不能: 刷屏时代如何摆脱行为上瘾 [M]. 阎佳, 译. 北京: 机械工业出版社, 2018.
- [11] GACKENBACH J. Psychology and the Internet: intrapersonal, interpersonal and transpersonal implications [M]. Academic Press, 1998.
- [12] CARBONELL X, PANOVA T. Acritical consideration of social networking sites' addiction potential [J]. *Addiction Research & Theory*, 2017, 25 (1): 48-57.
- [13] 毛峥, 姜永志. 神经质人格对问题性短视频使用的影响: 孤独感和无聊倾向的链式中介作用 [J]. 中国健康心理学杂志, 2022, 31 (3): 440-446.
- [14] 中国健康教育中心. 中国青少年健康教育核心信息及释义 (2018 版) [J]. 健康向导, 2018, 24 (6): 42-46.
- [15] BURNAY J, BILLIEUX J, BLAIRY S, et al. Which psychological factors influence Internet addiction? evidence through an integrative model [J]. *Computers in Human Behavior*, 2015, 43: 28-34.
- [16] KUSS D J, GRIFFITHS M D. Social networking sites and addiction: ten lessons learned [J]. *International Journal of Environmental Research and Public Health*, 2017, 14 (3): 311.
- [17] DAVIS R A. A cognitive-behavioral model of pathological Internet use [J]. *Computer in Human Behavior*, 2001, 17 (2): 187-195.
- [18] YOUNG K S. Psychology of computer use; XL. addictive use of the Internet: a case that breaks the stereotype [J]. *Psychological Reports*, 1996, 79 (3): 899-902.
- [19] 殷冬水. 法律滞后三论 [J]. 行政与法 (吉林省行政学院学报), 1998 (2): 28-30.
- [20] 齐学红. 关注青少年网络行为, 提高青少年信息素养 [EB/OL]. (2018-04-24). http://www.moe.gov.cn/jyb_xwfb/moe_2082/zl_2018n/2018_zl29/201804/t20180424_334086.html.
- [21] 中国网信网. 移动互联网未成年人模式建设指南 [EB/OL]. (2024-11-15). https://www.cac.gov.cn/2024-11/15/c_1733364304749288.htm.
- [22] KOH Y S. The Korean national policy for Internet addiction [M]. Springer International Publishing, Switzerland Press, 2015.
- [23] 中国互联网络信息中心. 第 5 次全国未成年人互联网使用情况调查报告 [EB/OL]. (2023-12-25). <https://www.cnnic.cn/n4/2023/1225/c116-10908.html>.
- [24] 贾开. 人工智能与算法治理研究 [J]. 中国行政管理, 2019 (1): 17-22.
- [25] 马长山. 算法治理的正义尺度 [J]. 学术前沿, 2022 (10): 68-76.
- [26] 李静辉. 算法推荐意识形态属性的生成逻辑、风险及治理 [J]. 理论导刊, 2022 (2): 70-76.
- [27] [德] 卡尔·拉伦茨. 德国民法通论 [M]. 王晓晔, 邵建东, 程建英, 等, 译. 北京: 法律出版社, 2013.

(下转第 92 页)

- [3] SCHUCH F B, VANCAMPFORT D. Physical activity, exercise, and mental disorders: it is time to move on [J]. Trends in Psychiatry and Psychotherapy, 2021, 43 (3): 177-184.
- [4] FOGG B J. Persuasive technology: using computers to change what we think and do [M]. San Francisco: Morgan Kaufmann Publisher, 2003.
- [5] ALDENAINI N, ALQAHTANI F, ORJI R, et al. Trends in persuasive technologies for physical activity and sedentary behavior: a systematic review [J]. Front in Artificial Intelligence, 2020; 3-7.
- [6] ALDENAINI N, ALSLAITY A, ORJI S R. Persuasive strategies and their implementations in mobile interventions for physical activity: a systematic review [J]. International Journal of Human-Computer Interaction, 2023, 39 (11/15): 2292-2338.
- [7] MICHIE S, VAN STRALEN M M, WEST R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions [J]. Implementation Science, 2011, 6: 42.
- [8] WENDEL S. 随心所欲: 为改变用户行为而设计 [M]. 张一弛, 孙锦龙, 译. 北京: 电子工业出版社, 2016.
- [9] OINAS-KUKKONEN H, HARJUMAA M. Persuasive systems design: key issues, process model, and system features [J]. Communications of the Association for Information Systems, 2009, 55 (9): 913-914.
- [10] HARTONO Y, MURNIATI M P. Goal setting theory: the effect of incentive moderation on individual performance [J]. Research in Management and Accounting, 2020, 3 (2), 95-106.
- [11] GARDINER C, BRYAN A. Randomized controlled trials of self-monitoring interventions with or without incentives for diet and exercise among individuals with overweight or obesity: psychological and behavioural effects [J]. British Journal of Health Psychology, 2021, 26 (4): 1114-1134.
- [12] 孙浩, 蔡宜静, 阎晋虎, 等. 社交媒体促进青少年体育行为的影响机制——基于混合方法的研究 [J]. 成都体育学院学报, 2024, 50 (5): 109-120.
- [13] MEMA E, SPAIN E S, MARTIN C K, et al. Social influences on physical activity for establishing criteria leading to exercise persistence [J]. PLoS One, 2022, 17 (10): e0274259.
- [14] 顾泽玉. 基于劝导式理论的运动健康类 APP 交互设计研究 [D]. 上海: 华东理工大学, 2021.
- [15] ORJI R, ALSLAITY A, CHAN G. Towards understanding the mechanism through which reward and punishment motivate or demotivate behaviours [J]. Behaviour & Information Technology, 2024, 43 (6): 1042-1066.
- [16] LIU N, YIN J, TAN S, et al. Mobile health applications for older adults: a systematic review of interface and persuasive feature design [J]. Journal of the American Medical Informatics Association, 2021, 28 (11): 2483-2501.
- [17] 刘欢欢. 基于健身 APP 的用户体验设计研究 [D]. 无锡: 江南大学, 2023.
- [18] 卢礼. 有氧运动中不同时间段音乐干预的情绪改善效果差异研究 [D]. 金华: 浙江师范大学, 2023.
- (收稿日期: 2025-06-25)
- 作者简介:**
孟瑶 (1980-), 通信作者, 女, 博士, 副教授, 主要研究方向: 人机交互、可穿戴计算。E-mail: yaomeng_elu@163.com。
鲁琴 (1971-), 女, 硕士, 教授, 主要研究方向: 软件开发、数据库。
祁春霞 (1979-), 女, 硕士, 讲师, 主要研究方向: 人工智能。

(上接第 85 页)

- [28] 李貌, 韩璞庚. 数字时代“信息茧房”束缚下主体性的解构与重建 [J]. 江苏社会科学, 2024 (3): 47-54.
- [29] 马长山. 迈向数字社会的法律 [M]. 北京: 法律出版社, 2021.
- [30] 张吉豫. 构建多元共治的算法治理体系 [J]. 法律科学 (西北政法大学学报), 2022, 40 (1): 115-123.
- [31] 丁道勤. 《电子商务法》平台责任“管道化”问题及其反思 [J]. 北京航空航天大学学报 (社会科学版), 2018, 31 (6): 1-6.
- [32] 张凌寒. 人工智能法律治理的路径拓展 [J]. 中国社会科学, 2025 (1): 91-110, 206.
- [33] 马长山. 数字司法的法治边界 [J]. 东方法学, 2024 (4): 127-142.
- [34] 翁跃强. 大数据分析在法律监督中的应用 [J]. 国家检察官学院学报, 2024, 32 (1): 98-113.
- [35] 王涛. 新质生产力驱动下的法治范式转型——以 DeepSeek 为例 [J]. 东方法学, 2025 (2): 137-152.
- (收稿日期: 2025-05-27)
- 作者简介:**
杨美玲 (2000-), 女, 硕士研究生, 主要研究方向: 网络法、民商法。
潘静 (1983-), 女, 博士, 副教授, 主要研究方向: 网络法、金融法。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.com