

生成式人工智能背景下网络舆情智能治理： 理论建构与实现路径

张 强, 王国硕

四川文化产业职业学院融媒学院, 四川 成都

收稿日期: 2026年7月23日; 录用日期: 2026年9月4日; 发布日期: 2026年9月11日

摘 要

生成式人工智能(AIGC)的快速迭代正深度重塑信息传播生态, 推动人类社会加速步入智能传播时代。作为影响网络社会运行和公共治理转型的重要变量, AIGC不仅改变了网络信息生产方式和传播机制, 也改变了网络舆情风险生成逻辑, 使传统网络舆情治理模式面临适应性挑战。本文基于智能治理、协同治理和风险治理等理论视角, 从现实逻辑、理论建构和实现路径三个层面对AIGC时代网络舆情智能治理展开分析。研究认为, 网络舆情智能治理并非人工智能技术与传统治理模式的简单结合, 而是在智能传播环境变化基础上形成的新型治理范式, 其理论框架体现为数据驱动、智能研判、人机协同和多元共治。推进网络舆情智能治理, 需要提升智能感知能力、强化智能研判能力、优化人机协同机制和完善多元协同机制, 并在伦理规制与制衡机制下, 实现技术赋能与治理价值的统一。

关键词

AIGC, 网络舆情, 智能治理, 人机协同, 风险治理

Intelligent Governance of Online Public Opinion in the Context of AIGC: Theoretical Construction and Implementation Pathways

Qiang Zhang, Guoshuo Wang

Media Convergence Institute, Sichuan Vocational College of Cultural Industries, Chengdu Sichuan

Received: July 23, 2026; accepted: September 4, 2026; published: September 11, 2026

Abstract

The rapid iteration of Artificial Intelligence Generated Content (AIGC) is profoundly reshaping the

文章引用: 张强, 王国硕. 生成式人工智能背景下网络舆情智能治理: 理论建构与实现路径[J]. 社会科学前沿, 2026, 15(9): 182-190. DOI: 10.12677/ass.2026.159742

information communication ecosystem, accelerating humanity's entry into the era of intelligent communication. As a critical variable influencing the operation of online society and the transformation of public governance, AIGC has not only transformed the modes of online information production and dissemination mechanisms but also fundamentally altered the generative logic of online public opinion risks, subjecting traditional governance models to adaptive challenges. Drawing upon the theoretical perspectives of intelligent governance, collaborative governance, and risk governance, this paper conducts a systematic analysis of intelligent governance of online public opinion in the AIGC era across three dimensions: practical logic, theoretical construction, and implementation pathways. The study posits that intelligent governance of online public opinion is not a mere combination of artificial intelligence technologies with traditional governance models, but rather a novel governance paradigm emerging from the transformative changes in the intelligent communication environment. Its theoretical framework is embodied in four pillars: data-driven decision-making, intelligent assessment, human-machine collaboration, and pluralistic co-governance. Advancing intelligent governance of online public opinion requires enhancing intelligent perception capabilities, strengthening intelligent assessment capabilities, optimizing human-machine collaboration mechanisms, and improving pluralistic coordination mechanisms, while under the framework of ethical regulation and checks-and-balances mechanisms, achieving the integration of technological empowerment and governance values.

Keywords

AIGC, Online Public Opinion, Intelligent Governance, Human-Machine Collaboration, Risk Governance

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. 引言

生成式人工智能(AIGC)的快速发展正在推动数字传播生态发生深刻变革,并成为影响网络社会运行和公共治理转型的重要变量。不同于早期人工智能主要承担信息处理和辅助分析功能,AIGC驱动的智能传播革命,正在从底层重构信息生产、分发和消费的全链条,使网络信息生成主体、传播机制和风险形态发生系统性变化,这标志着世界已迈入以人机双智能协同为特征的智能传播时代[1]。

在此背景下,网络舆情的性质已发生根本变化:它不再仅仅是社会事件在网络空间中的信息映射,而是逐渐演化为技术系统、平台架构、用户行为与社会情绪多重力量交互作用的动态产物。与此同时,建立在搜索引擎、经验判断和行政处理等基础上的早期舆情治理体系,面临着治理对象复杂化、风险识别滞后化和响应机制适配不足等新挑战。如何理解 AIGC 对网络舆情运行逻辑的重塑,并推动治理体系实现智能化转型,成为网络治理现代化进程中亟待回应的重要议题。

关于智能传播时代的治理问题,国内外学者形成了三条重要的研究脉络。在平台治理领域,Robert Gorwa 提出了平台治理三角模型[2];国内学者张志安认为,平台算法治理实质上是“空间治理”,难以单纯依靠技术层面的调整来解决全民性问题[3]。在算法治理领域,John Danaher 提出“算法统治”概念,警示算法决策的不可解释性对公共决策合法性的侵蚀[4];沈浩则强调应从社会主流价值导向、用户权益保护、监管体系完善和企业社会责任强化等维度构建算法治理的支撑体系[5]。在 AI 伦理治理领域,陈禹含等提出 AI 决策权应掌握在人类手中,保持人类主体的判断力,确保技术的合规性[6];高奇琦提出必须建立涵盖内容生成端、平台传播端和用户接收端的全链条治理机制[7]。上述研究为理解 AIGC 时代的舆

情治理提供了重要参照,但既有成果多聚焦于平台作为内容中介的规制或一般性算法治理议题,对生成式 AI 大规模介入内容生产后治理逻辑的根本性变化缺乏系统回应。当聚焦到人工智能赋能网络舆情治理研究时,目前涉及网络舆情智能治理的文献基本围绕其必要性、可行性和实现路径进行讨论[8][9]。遗憾的是,这些文献竟未捋清网络舆情智能治理的概念和特征等基本属性。总体而言,学界对 AIGC 时代网络舆情智能治理的系统性理论建构仍显薄弱。基于此,本文引入智能治理、协同治理和风险治理理论,进一步分析网络舆情治理转型的现实逻辑,建构网络舆情智能治理的理论框架,并从智能感知、可信研判、协同响应和制度保障等方面探讨其实现路径,以期智能传播时代网络舆情治理现代化提供理论参考。

2. 生成式人工智能背景下网络舆情治理的现实逻辑

2.1. 网络舆情传播生态的智能化重构

AIGC 的广泛应用正在推动网络传播生态发生结构性变化,并成为网络舆情治理转型的重要现实基础。与传统信息生产模式相比,AIGC 突破了人工内容生产在技术门槛、时间成本与专业能力等方面的限制,使网络信息生产由人工主导逐步转向人机协同,传播生态由此呈现全新的运行特征。

首先,AIGC 推动网络信息生产主体更加多元化。传统网络信息生产主要依赖用户、媒体机构和组织化主体完成,而生成式人工智能大幅降低了内容创作门槛,使普通用户与社会组织能够借助智能工具快速生成文本、图像与音视频等多模态内容。信息生产能力的泛化在扩大网络表达空间的同时,也使信息来源更趋复杂,显著增加了内容真实性甄别与风险识别的难度。其次,AIGC 推动网络传播机制从社交扩散向算法驱动传播转变。在传统传播格局中,信息扩散高度依赖用户互动与社会关系网络,而在智能传播环境下,平台推荐算法、智能分发机制与生成内容深度耦合,使信息流动更加依赖算法排序与平台规则,信息传播的走向不再完全由用户主动选择所决定。再次,AIGC 深刻改变了网络舆情风险的发生机理。生成式人工智能大幅降低了虚假信息、误导性内容与情绪化表达的制造成本,同时算法推荐机制可能进一步放大具有争议性与情绪唤醒效应的信息,推动舆情风险从传统的事件触发模式,转向智能生成、算法扩散与社会互动协同驱动的复合型模式。上述三重变化表明,AIGC 时代网络舆情治理的指向已不再局限于信息内容本身,而是扩展为涵盖内容生产、传播机制与风险演化全链条的复杂传播生态系统。

2.2. 网络舆情治理环境的系统性变迁

如果说传播生态的变化回答了“什么变了”的问题,那么治理环境的变迁则进一步追问:这些变化给既有治理体系带来了怎样的约束与压力。

传统网络舆情治理主要依赖人工监测、经验研判与事后处置[10],其有效性建立在信息来源相对稳定、传播路径较为清晰、风险演化具有一定规律性等前提之上。但在智能传播环境下,上述前提正在逐一瓦解。其一,信息来源可追溯性的消解。人工智能参与内容生产后,机器生成内容与人工生成内容深度交织,传统基于来源可信度的判断标准因此失效,内容溯源与责任认定变得异常困难。其二,风险演化速度与治理响应速度之间的落差急剧扩大。算法推荐机制从根本上改写了舆情传播的底层逻辑,风险信息可能在平台算法、用户互动与智能生成内容的多重叠加作用下呈指数级扩散,传统以人工搜索与热点追踪为主的信息获取方式,难以在风险萌芽阶段实现有效识别。其三,治理任务边界持续扩展。人工智能嵌入信息生产与传播全过程后,算法透明性、平台责任归属、数据安全防护与技术伦理规范等问题相继浮出水面,网络舆情治理不再仅仅是对信息内容的单向管理,而是涉及技术系统、平台机制与社会关系网络的复合治理过程。

因此,这些变化意味着传统治理模式所依赖的初始条件已发生根本性改变,单纯在既有框架内修补工具已不足以回应现实挑战。

2.3. 网络舆情智能治理的生成逻辑

传播生态的重构与治理环境的变迁, 共同构成了治理范式转型的倒逼机制。这种转型并非单纯的技术升级, 而是传播生态的结构性变迁与治理环境的系统性压力叠加作用下的必然结果。

一方面, 人工智能、大数据与算法技术的迅速发展, 为网络舆情风险的识别、趋势预判与辅助决策提供了前所未有的能力支撑, 使被动响应向主动预警的转变具备了技术可行性。另一方面, 智能治理强调技术能力与制度安排的有机融合——其要义不在于以技术逻辑替代治理逻辑, 而在于借助人工智能的辅助功能, 推动治理主体、技术系统与社会力量之间形成新型协同关系。技术提供了可能性, 环境变化则提供了必要性, 二者共同构成了智能治理生成的完整驱动力。

因此, AIGC 时代网络舆情智能治理本质上体现为一种治理范式的转换。其核心在于以智能技术提升治理效能, 同时以制度规范、主体协同与价值引导确保治理方向不发生偏移, 从而实现技术效率与公共价值的有机统一。从传统治理向智能治理的跃迁, 既是对智能传播环境变化的现实回应, 也是推动网络治理体系现代化的重要路径选择。

3. 生成式人工智能背景下网络舆情智能治理的理论建构

3.1. 网络舆情智能治理的理论来源

网络舆情智能治理并非单纯由技术发展推动, 而是融合了智能治理、协同治理和风险治理等多种理论资源。这些理论分别从技术赋能、主体关系和风险管理三个维度, 为理解 AIGC 时代网络舆情治理提供理论支撑。

智能治理理论为网络舆情智能治理提供技术赋能视角。该理论强调, 现代公共治理需要充分利用数字技术和智能系统提升信息处理、资源配置和决策支持能力。与传统数字治理主要关注信息化建设不同, 智能治理更加关注人工智能在复杂环境中的分析和预测能力。在 AIGC 时代, 人工智能已经由辅助性工具逐渐参与内容生成、信息传播和风险识别过程, 因此网络舆情治理需要重新认识技术在治理体系中的功能定位。

协同治理理论为网络舆情智能治理提供主体关系解释。网络舆情治理涉及政府部门、网络平台、媒体组织、技术企业以及公众等多元主体^[11]。随着 AIGC 降低内容生产门槛, 信息传播主体进一步多元化, 单一治理主体难以独立应对复杂舆情风险。协同治理强调不同主体之间通过责任共享、资源整合和互动合作形成治理合力, 为构建开放协同的网络舆情治理机制提供理论依据。

风险治理理论则为网络舆情智能治理提供全过程治理视角。传统网络舆情治理更多关注事件发生后的处置, 而风险治理强调风险识别、评估、预警和干预全过程管理。AIGC 时代, 网络舆情风险可能在内容生成、算法推荐和用户互动过程中持续演化, 因此需要由被动响应转向主动预防。

由此, 智能治理理论回答了“如何利用技术提升治理能力”的问题, 协同治理理论回答了“由谁共同参与治理”的问题, 风险治理理论回答了“如何实现全过程风险控制”的问题。三种理论共同构成网络舆情智能治理的理论基础。

3.2. 网络舆情智能治理的概念内涵

前述三种理论分别从技术赋能、主体关系与风险管理三个维度回应了网络舆情智能治理的核心问题, 共同构成了其学理基础。基于上述理论资源, 可将网络舆情智能治理界定为: 在 AIGC 深度介入网络信息生产与传播过程背景下形成的新型治理模式, 是指运用人工智能、大数据、算法分析等智能技术, 对网络舆情信息进行动态感知、综合研判和协同响应, 并通过多主体参与实现网络舆情风险有效治理的过程。

从常见的概念关系来看, 网络舆情智能治理与传统网络舆情治理、智慧治理和算法治理既存在联系,

也具有区别。与传统网络舆情治理相比, 智能治理更加突出数据驱动和智能辅助, 强调利用人工智能提升治理全过程能力; 与智慧治理相比, 智能治理更加关注人工智能技术如何进入具体治理过程, 通过增强感知、分析和决策能力推动治理方式变化; 与算法治理相比, 智能治理并非主要关注算法规则本身, 而是强调如何利用智能技术服务治理目标, 实现技术能力与公共价值的统一。因此, 网络舆情智能治理的本质不是技术治理化, 而是治理智能化。其核心在于通过人工智能赋能网络舆情治理体系, 使治理主体具备更加主动、精准和协同的风险应对能力。

3.3. 网络舆情智能治理的基本特征

基于上述概念界定, 网络舆情智能治理作为智能传播环境下形成的新型治理范式, 其基本特征体现为数据驱动、智能研判、人机协同和多元共治四个方面。四个特征并非相互独立的技术要素, 而是共同构成网络舆情智能治理由信息感知、风险认知到协同治理的运行逻辑。

首先, 数据驱动构成网络舆情智能治理的基础。AIGC 时代, 网络信息生产规模持续扩大, 舆情治理面对的数据来源更加丰富、信息类型更加多样, 需要依托多源数据获取、整合和分析能力, 实现对网络信息变化和风险线索的动态感知。数据不仅是智能治理的重要资源, 也是连接风险识别、趋势分析和治理决策的重要基础。通过数据驱动, 网络舆情治理能够突破传统经验判断局限, 提高治理过程的主动性和精准性。

其次, 智能研判体现网络舆情智能治理的核心能力。网络舆情的生成与演化, 本质上交织着信息流动、情绪共振、利益博弈与价值认同等多重逻辑, 远非单一的信息扩散过程所能涵盖。面对复杂舆情事件, 简单的频次统计与热度排序已无法穿透舆情的深层结构, 必须借助大语言模型、知识图谱与智能分析技术, 对舆情要素进行跨维度整合与深度关联分析, 为治理主体提供超越经验直觉的判断支撑。

再次, 人机协同体现网络舆情智能治理的运行方式。人工智能具有快速处理信息和辅助分析决策的优势, 但网络舆情治理同时涉及价值判断、社会沟通和公共决策, 仍然需要人的参与。因此, 智能治理强调人工智能辅助与人的判断能力相结合, 通过明确技术系统与治理主体的功能边界, 实现技术效率与治理经验优势互补。

最后, 多元共治体现网络舆情智能治理的制度要求。网络舆情治理已经超越单一主体管理范畴, 需要政府、平台、媒体、技术企业和公众共同参与。通过责任协同、资源共享和利益协调, 可以形成更加开放、多元和持续的治理体系。

总体而言, 数据驱动提供治理基础, 智能研判提升认知能力, 人机协同优化治理过程, 多元共治保障治理运行, 共同构成 AIGC 时代网络舆情智能治理的理论框架。

4. 生成式人工智能背景下网络舆情智能治理的实现路径

AIGC 时代网络舆情智能治理的实现, 需要以智能技术赋能为基础, 以治理机制优化为保障, 推动网络舆情治理由被动响应向主动治理转变。智能治理并不是单纯提高技术应用水平, 而是围绕风险感知、分析研判、协同决策和制度保障形成系统化治理能力。因此, 需要从智能感知、智能研判、人机协同和多元协同等方面推进治理体系优化。

4.1. 提升智能感知能力, 实现网络舆情风险主动识别

智能感知是网络舆情智能治理的基础环节, 其核心在于利用人工智能、大数据和多模态分析技术, 提高对网络信息变化和潜在风险的发现能力, 实现治理关口前移。

传统网络舆情监测主要依赖人工搜索、关键词预警^[12]以及热点追踪, 能够发现已经形成影响的舆情

事件,但对于早期风险信号和隐性情绪变化识别能力有限。2025年7月起,山东烟台发生的一起“AI水军”抹黑新能源汽车案集中暴露了这一短板(以下简称山东烟台“AI水军”案)。该案中,理想、鸿蒙智行、小米等企业相继报案,称某网络平台出现大量涉及相关品牌新能源汽车的负面文章。警方查明,这是一个利用AI技术批量制造负面信息的新型网络“水军”团伙——通过AI软件对抓取的热点词汇和内容进行重组、改写,伪装成个人体验批量发布,实现“零成本”生产。短短数月内,仅理想汽车订单即锐减5000余辆,总市值下降约200亿元[13]。传统依赖关键词密度和情感极性的监测方式,面对AI批量生成、表述各异的海量“黑稿”,既无法识别内容的AI生成特征,也难以判断信息的真实性——这正是智能感知能力应前置介入的场景。在AIGC环境下,网络信息生产规模持续扩大,文本、图像和音视频等多模态内容交织交融传播,单一的信息监测方式难以满足复杂传播环境需求。因此,应推动网络舆情感知方式由信息采集向智能识别转变。一方面,利用大语言模型、多模态分析和数据融合技术,实现对不同类型信息的综合识别,提高风险发现效率;另一方面,加强数据关联分析,将信息变化、传播路径、用户情绪和社会思潮等因素结合起来,提高对舆情演化趋势的判断能力。

该案中,警方正是通过研判发现“发布账号数以千计,IP地址集中、发文时间集中”等异常现象才锁定目标。这揭示了一个关键方向:智能感知系统不应止于关键词匹配,而应通过账号行为分析识别批量发布的时间集中性、IP集中性等异常信号,在舆情大规模扩散前发出预警。智能感知的深层目标不止于捕捉信息层面的异常波动,更在于对海量异构内容进行语义特征提取与情感极性判别,从中识别具有舆情演化潜力的风险信号。为此,须推动人工智能与领域知识库、舆情知识图谱及社会语境数据深度融合,使智能系统在识别信息内容的同时,具备对社会意义场域的解析能力。通过智能感知能力的持续提升,网络舆情治理可从风险发现迈向风险预警,为后续研判与响应奠定扎实基础。

4.2. 强化智能研判能力,实现网络舆情风险精准认知

研判工作是网络舆情工作过程的重要内容,也是网络舆情工作成功与否的先决因素[14]。智能研判是网络舆情智能治理由风险发现走向风险认知的重要环节,其核心在于利用人工智能技术提升复杂舆情分析和趋势预测能力。

AIGC时代网络舆情形成机制更加复杂,单纯依靠数据统计难以准确理解事件发展逻辑。网络舆情不仅涉及信息数量变化,还包含主体关系、情绪变化、社会影响和价值诉求等多重因素。因此,智能研判需要突破简单的信息分析模式,提升对复杂社会情境的综合理解能力。山东烟台“AI水军”案中,“表述各异但指向相同”的海量信息——数千个账号发布不同措辞但同样指向同一品牌质量问题的负面内容——正需要借助大语言模型进行语义聚类 and 关联分析,将碎片化的负面信息整合为有组织的舆情攻击图谱。然而,AIGC生成内容仿真度高、逻辑自洽,仅凭文本内容难以判断其真实性,传统研判模式既缺乏对内容生成来源的追溯能力,也缺乏对批量发布行为的关联分析能力。因此,在技术操作层面,应以大语言模型为底层语义引擎,嵌入领域知识图谱与政策案例库,以提升模型对专业语境、历史参照和社会关系网络的理解能力。同时,应加强多源信息关联分析,通过语义理解、趋势预测和风险评估等方式,提高对舆情发展方向的判断能力。但需要强调的是,智能研判并不意味着将决策完全交由算法完成,而是在人工智能辅助基础上,由治理主体结合现实情境进行综合判断。通过强化智能研判能力,可以推动网络舆情治理由经验判断向智能辅助判断转变,提高风险分析的科学性和有效性。

4.3. 优化人机协同机制,实现网络舆情治理效能提升

人机协同是网络舆情智能治理区别于传统治理模式的重要特征,也是实现技术能力与治理能力有效结合的关键机制。

人工智能具有信息处理速度快、数据分析能力强等优势,能够承担信息筛选、风险识别、趋势预测

和辅助决策等任务,有效提升网络舆情治理效率。但网络舆情治理并非单纯的信息处理过程,而是涉及价值判断、社会沟通、利益协调和公共决策的复杂治理活动,人工智能无法完全替代人的主体作用。因此,智能治理需要明确人工智能与治理主体之间的功能边界。人工智能主要承担信息分析、风险提示和决策辅助等任务,治理主体则负责价值判断、政策把握和责任承担。人工智能的发展目标不是替代人类,而是增强人的认知和决策能力[15]。通过合理分工,可以实现技术优势与治理经验的互补,提升网络舆情治理的整体效能。

山东烟台“AI水军”案中, AI批量生成“黑稿”的能力远超人工审核效率——该团伙注册多家公司、操纵数千个平台账号,发布海量内容,仅凭人工几乎不可能逐一甄别。应建立“AI识别AI”的辅助审核机制:由智能系统完成海量内容的初筛和标记,标注疑似AI生成、疑似批量发布等风险信号,再由人工审核员进行重点研判和处置。警方最终通过人工侦查查明团伙的组织架构和牟利模式——该团伙不敲诈、不引流,仅依靠大量账号发文的流量收益牟利,近年来收到平台打款50余笔共计约180万元,抓获犯罪嫌疑人12名,关停违法账号8000余个[13]。这一过程表明:智能系统可以完成初筛和预警,但团伙行为的定性、取证和法律适用仍需人类专业判断。在实践过程中,应建立智能辅助与人工决策相结合的协同机制。一方面,加强智能系统在信息处理、趋势分析和风险预警中的应用,提高治理效率;另一方面,加强专家研判、人工审核和责任监督,避免算法偏差、数据局限或模型误判影响治理效果。

同时,应避免将智能治理简单理解为技术替代治理。网络舆情治理具有明显的社会属性,技术系统需要在制度规范和价值目标引导下运行。只有实现人工智能技术能力与治理主体实践能力的有效结合,才能真正推动网络舆情治理由传统人工处置向智能辅助、多方协同转变。

4.4. 完善多元协同机制, 实现智能治理规范运行

网络舆情智能治理不仅需要技术能力支撑,也需要多元主体协同机制保障。AIGC提高网络舆情治理效率的同时,也带来了算法透明、数据安全、责任划分和技术伦理等新的治理问题。因此,需要推动技术应用、主体协同和制度规范相互结合,实现智能治理规范运行。

首先,应完善算法治理机制,提高智能系统运行的透明度和可解释性。网络舆情智能治理依赖算法模型进行信息识别和风险分析,而算法偏差可能影响治理判断和资源配置。因此,需要加强算法评估、运行监督和结果审查,确保智能系统运行符合公共治理目标。其次,应强化平台主体责任,推动多主体协同治理。网络平台既是信息传播空间,也是算法推荐和内容分发的重要参与者,在网络舆情风险形成和扩散过程中具有重要影响。数字平台已经成为现代传播秩序的重要组织者,需要承担相应的规则制定和责任保障功能[16]。山东烟台“AI水军”案中,该团伙注册多家公司、操纵数千个平台账号,仅依靠大量账号发文的流量收益牟利——这一作案模式暴露了平台在账号注册审核、内容批量发布监测、AI生成内容识别等方面的治理漏洞。因此,应推动平台承担内容管理、风险预警和算法治理责任,同时促进政府、平台、媒体、技术企业和公众之间形成协同关系。各个主体各司其职,协同合作,共同构建网络舆情监管的多元共治格局[17]。再次,应加强数据安全和伦理规范建设,确保人工智能应用在合法、安全和可控范围内运行。智能治理不能仅关注技术效率,还需要关注公共价值、社会公平和责任边界,通过制度建设提升智能治理的可持续性。

通过完善多元协同机制,可以推动网络舆情智能治理由单一技术应用走向系统治理,实现技术能力、治理机制和公共价值之间的协调统一。

5. 生成式人工智能背景下智能治理的伦理规制与制衡机制

前述的智能感知、智能研判、人机协同与多元共治四条路径,旨在从技术能力与组织架构层面提升

治理效能。然而, 效能的提升与技术伦理风险之间始终存在内在张力。AIGC 深度嵌入舆情治理全过程后, 治理体系自身的结构性缺陷开始暴露: 算法训练数据的偏差可能被系统放大, 形成歧视性循环; “风险识别”名下的数据采集与深度分析, 容易滑向泛监控化; 监控能力的提升不意味着监控边界的无限扩展, 须防止“预警”异化为“常态监控”; 核心技术集中于少数科技企业, 政府治理日益依赖外部技术供给, 可能形成隐性权力转移, 削弱公共治理的自主性。

上述风险源于治理体系内部, 而外部技术环境的恶化则进一步放大了这些挑战。山东烟台“AI 水军”案提供了警示性样本: AI 工具被低成本地滥用于制造虚假舆情, 而现行治理体系对“AI 辅助造谣”的识别、溯源和责任认定能力严重不足——这既是前文所述识别滞后、研判失效等结构性短板的集中体现, 也反过来加剧了治理体系的伦理风险。当治理技术本身存在缺陷, 而外部威胁又在持续升级时, 技术赋能与技术滥用之间便形成了一种危险的对称。正如何小英等所言, 算法推荐承载着多重价值意涵^[18], 但客观表象下往往掩盖着结构性的不公。

针对上述内外部交织的多重风险, 须同步嵌入三重制衡机制。其一, 算法审计机制。建立常态化的影响评估与审计制度, 对数据输入、模型设计与输出结果进行定期审查, 检测并纠正系统性偏见, 确保算法在关键决策环节具备可追溯、可审查的技术条件。其二, 公众参与机制。构建多元主体参与的伦理监督平台, 在系统部署与运行中设置公开征求意见环节, 建立独立伦理咨询委员会, 定期发布运行报告并接受社会评议, 避免治理规则由单一主体单方面决定。其三, 申诉与救济机制。畅通公民申诉渠道, 明确智能辅助决策的责任归属与问责路径, 建立人工复核机制, 确保受决策影响的个人与群体享有知情权、异议权与救济权。

综上, 技术赋能与伦理制衡构成智能治理的一体两面。技术与权利的平衡不能仅依赖系统自我优化, 须通过制度化的制衡机制予以保障。唯有在提升治理效能的同时建立起坚实的制衡框架, 方能实现治理效能与公共价值的辩证统一。

6. 结语

生成式人工智能的发展推动网络传播生态和舆情风险形态发生深刻变化, 网络舆情治理正经历由传统经验模式向智能治理范式的系统性转变。AIGC 深度嵌入信息生产与传播链条后, 舆情风险在内容生成、算法分发和社会互动三个环节叠加演化, 传统依赖人工监测和事后响应的工作机制, 在面对智能传播环境下信息规模爆炸性增长与风险形态持续变异时, 暴露出识别滞后、研判不足和响应乏力等结构性短板, 使智能化转型成为了现实必需。山东烟台“AI 水军”案便是上述困境的集中体现——AI 批量生成的“黑稿”在短时间内造成企业市值蒸发约 200 亿元, 而传统治理模式在识别、研判与响应三个环节均暴露出结构性失灵, 充分印证了治理范式转型的紧迫性与现实必要性。

针对网络舆情的智能治理, 本文从现实逻辑、理论建构与实现路径三个维度展开了分析, 尝试回应 AIGC 时代网络舆情治理“为何转、依何治、如何行”的递进式命题。在现实逻辑层面, 研究揭示了传播生态重构、治理环境变迁与智能治理生成之间的内在关联, 指出技术变量已从治理外部因素转化为治理内部结构要素。在理论建构层面, 整合智能治理、协同治理与风险治理三种理论资源, 界定了网络舆情智能治理的概念内涵, 提出数据驱动、智能研判、人机协同与多元共治四位一体的理论框架, 辨析了四者之间的功能关联。在实现路径层面, 研究提出以提升智能感知能力实现关口前移, 以强化智能研判能力推动认知升级, 以优化人机协同机制释放互补效能, 以完善多元协同机制构建制度基底, 四条路径相互支撑, 共同构成从技术嵌入到制度内化的转化通道。

值得强调的是, 智能治理的技术工具属性与公共价值属性始终构成一体两面。算法效率不能替代治理判断, 数据预测不能覆盖社会情境, 技术赋能不能消解制度约束。网络舆情智能治理的深层命题, 不

在于人工智能能够做什么, 而在于治理体系应当如何规约技术应用的方向与边界。唯有在技术创新与制度规范之间维持动态平衡, 在效率追求与价值坚守之间建立协调机制, 方能推动智能技术从治理工具真正升维为治理范式, 为网络空间治理现代化提供可持续的内生动力。

基金项目

四川省首批哲学社会科学重点实验室——数智公共治理实验室、四川省哲学社会科学重点研究基地——区域公共管理信息化研究中心 2025 年度资助项目“生成式人工智能背景下网络舆情智能治理的路径与机制研究”阶段性研究成果(课题编号为 QGXH25-06)。

参考文献

- [1] 黄一涛, 杜友君. 智能传播时代我国游戏出版业高质量发展研究[J]. 理论月刊, 2025(11): 107-119+159-160.
- [2] Gorwa, R. (2019) The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content. *Internet Policy Review*, 8, 1-22. <https://doi.org/10.14763/2019.2.1407>
- [3] 算法投射真实社会问题, 学者建议引入“空间治理”思路[EB/OL]. <https://www.jfdaily.com.cn/statics/res/html/web/newsDetail.html?id=1040929>, 2025-12-24.
- [4] Danaher, J. (2016) The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy & Technology*, 29, 245-268. <https://doi.org/10.1007/s13347-015-0211-1>
- [5] 沈浩. 构建算法治理落地支撑体系[N]. 人民邮电, 2022-06-07(005).
- [6] 陈禹含, 邓吉阵, 宋信强. 重构“艺术真实”: AIGC 时代纪录片影像美学的嬗变与创新[J]. 视听, 2026(15): 64-68.
- [7] 高奇琦. GPT 技术与国家治理现代化——基于秩序、赋权与创新的框架[J]. 山东大学学报(哲学社会科学版), 2023(5): 147-156.
- [8] 徐芸怡. 智能媒体时代网络舆情的智能治理[J]. 新媒体研究, 2021, 7(9): 5-7+45.
- [9] 方堃, 李帆. 涉民族因素网络舆情的智能治理: 逻辑与路径[J]. 黑龙江民族丛刊, 2021(2): 34-38.
- [10] 陈祥富. 生成式 AI 赋能网络舆情智能治理: 应用路径与效能提升研究[J]. 传播与版权, 2026(6): 108-111.
- [11] 高崧韬, 刘峻宏, 宫梦莹. 大数据环境下多元主体协同的网络舆情治理能力研究[J]. 情报科学, 2023, 41(12): 41-47.
- [12] 戴永红, 莫一帆, 付乐. “一带一路”沿线国家中资企业“AI + ESG”能力建设与提升机制[J]. 当代中国与世界, 2025(4): 85-95.
- [13] 王阳. 利用 AI 抓热点, 炮制“爆款”博流量: 一起底新型网络“水军”犯罪手法[N]. 新华每日电讯, 2026-04-09(006).
- [14] 魏敏. 基于信息共享的网络舆情信息工作机制建构[J]. 新闻研究导刊, 2017, 8(20): 21-23.
- [15] Daugherty, P.R. and Wilson, H.J. (2018) Human + Machine: Reimagining Work in the Age of AI. Harvard Business Review Press.
- [16] Van Dijck, J., Poell, T. and De Waal, M. (2018) The Platform Society. Oxford University Press. <https://doi.org/10.1093/oso/9780190889760.001.0001>
- [17] 倪炜宏. 新媒体环境下浙江省乐清市网络舆情政府监管问题研究[D]: [硕士学位论文]. 乌鲁木齐: 新疆农业大学, 2025.
- [18] 何小英, 潘永康. 数字时代下算法推荐权力的表征形式及成因分析[J]. 重庆邮电大学学报(社会科学版), 2025, 37(6): 150-158.