ISSN：3105-0085
EISSN：3105-0093

从系统演进到智能挑战：生成式人工智能时代的网络社会治理研究

周行虎

重庆师范大学，重庆 401331

摘　要：我国网络社会治理经历了从技术准备，到框架搭建再到深化拓展的持续过程，其动力既来自互联网发展现状，也来自党和政府对时代与国情的把握。随着时代的发展，我国网络社会治理的理念逐渐完成由管理到治理的转向，应对方式也从被动反应转为主动作为，治理手段变单一管理为综合治理。但人工智能时代的到来，让网络空间发生了变革，生成式人工智能一方面在信息快速传播方面发挥了巨大作用，另一方面也带来了虚假信息、深度伪造和版权边界模糊，以及平台责任失衡等问题。这一过程显示出网络社会的复杂性，因此对其的治理应不断更新方式手段，以制度上的创新回应社会治理需求。

关键词：网络；社会治理；生成式人工智能

DOI：10.64649/yh.shygl.issn3105-0085.202610009

1 概念界定

网络社会治理是指在已有的社会治理规则基础上，政府、企业和个人等多元组织，针对互联网相关事物，做出的治理实践活动。目前其已成为国家治理体系中的重要组成部分，学界也在不断完善网络社会治理的内涵。王芳指出，网络社会是由互联网联系起来的跨越国界的复杂巨系统，其治理需探索与之相适应的新型治理结构、方式与机制[1]。她认为网络社会治理应当整合治理环境、主体、对象、目标、策略、技术、效果等多种要素，从而构建一个系统化的理论模型。徐汉名与王玉梅则从构建网络社会治理法学出发，将其研究范围拓展到网络安全、核心技术、个人信息保护、网络交易与网络行为五大法律规范体系[2]。此外，网络社会自身特点也要求治理模式迭代更新，从而更好地适应外部变化[3]。如今生成式人工智能的发展，进一步拓展了治理边界，网络治理内涵也在动态改变。

2 发展历程

我国网络社会治理主要随着互联网技术和改革开放的变化而变化，不同阶段各有特征，但总有一个核心议题存在。根据目前的文献，能把历程大致划分为四个阶段[4]。

2.1 技术储备与体制奠基

1978 年党的十一届三中全会的召开，标志着中国进入改革开放的新时期，这个阶段虽然还没有正式接入互联网，但为网络时代的到来做了基础设施与人才的准备。，邓小平同志提出的“开发信息资源，服务四化建设”成为指导方针，信息资源得到了重视。该阶段的核心重点就是抓住信息革命的机遇，为后续发展奠定技术和制度基础[5]。

2.2 全功能接入与依法管理起步

中国在 1994 年 4 月 20 日接入国际互联网，正式开启了互联网时代。在刚接入时，互联网的双刃剑效应就已经开始显现，这体现在它既带来了经济和社会效益，也让网络色情、暴力、诈骗等违法行为层出不穷。在这个背景下，政府成为治理的主体，主要从两个方面展开行动：第一，建立健全互联网制度体系，推进依法管理。例如，2000 年《全国人民代表大会常务委员会关于维护互联网安全的决定》通过，正式开始将互联网管理纳入法制建设中。第二，不断完善互联网基础设施，开始多元治理。如 1997 年中国互联网络信息中心成立，实现了国内四大骨干网的互联互通。同时宣传部等部门在治理中的地位得到提高，管理体制也在不断完善。2001 年 5 月，中国互联网协会成立，行业自律与公众监督并举，显示出社会力量开始参与治理中。

2.3 内容监管成为核心议题

这个阶段我国互联网高速发展，产业蓬勃发展，网民规模在 2008 年扩大到 2.53 亿。而 Web2.0 时代的到来，让网络空间舆论更加复杂多元，历史虚无主义、民粹主义、新自由主义等各种思潮在互联网中传播，冲击了主流意识形态，该阶段突出表现为网络意识形态斗争日益尖锐化。党和政府在把握时代变化情况下，开始主导互联网治理进入坚持主流意识形态的新阶段[6]。这个阶段的治理重心也转变为内容监管和舆论引导，目的是既把互联网作为一个治理的创新工具，同时又确保它沿着正确方向发展。

ISSN：3105-0085
EISSN：3105-0093

2.4 国家战略与综合治理体系成型

党的十八大以来，中国进入新时代，互联网环境发生了更深刻的变化。在一些网络安全事件后，党和国家更加重视网络安全，开始将其上升到国家战略层面，该阶段的核心也就发展成了维护国家安全。在这一阶段，治理活动主要围绕明确建设网络强国的发展目标、从维护国家安全和稳定的战略高度系统部署网络安全工作、积极推动互联网全球共治共享三个方面展开。治理体系也开始变革，在领导机制上，2014 年成立中央中央网络安全和信息化领导小组，并在 2018 年升格为中央网络安全和信息化委员会；在法治建设上，2017 年《中华人民共和国网络安全法》正式施行，此后又出台了《数据安全法》和《个人信息保护法》；在治理范围上，从早期的基础设施和内容管理，扩展到数据安全、算法治理、未成年人保护、人工智能伦理等新兴前沿领域 [7]。通过一系列的设计和布局，逐渐形成了“党委领导、政府管理、企业履责、社会监督、网民自律”的社会协同共治格局。

3 治理新挑战

3.1 虚假信息与深度伪造的双重压力

现如今生成式人工智能技术发展迅速，在为社会带来便捷的同时，也造成了更容易制造虚假信息和深度伪造的局面，网络社会治理格局需要适时调整。[8] 一项对主流 AI 的测试显示，给它一张普通照片和简单的文字指令，就能获得一张符合自己要求的假图。更进一步，若不断调整对话，还可以对生成图进行细节调整，让人防不胜防。还有一些软件特别提供电商售后模板，专门帮助生成假的商品图，从而规避平台审核。

与之相比，对 AI 生成内容的识别技术还相对落后，现有检测工具准确率不高，特别是在对二次创作的深度伪造图的识别上，还有很大的进步空间。这一方面是因为现有技术更新迭代快，生成能力不断增强，已经和真实事物差别很小；另一方面则在于伪造者只需要略微截图调整，就能隐藏使用 AI 的痕迹。当然第三方专业鉴定的高昂成本也是重要原因，中小商家最终会迫于成本，放弃追究，这也一定程度助长了造假行为。

3.2 版权秩序的重构

生成式人工智能所涉及的版权问题源于它要依赖大量数据，而开发者若想使模型有足够强的性能就必须持续地获得大体量高质量数据。就目前所用的开放资源、购买的数据库及爬虫方法来说，有关信息的获取仍不能完全避开著作权保护期尚未届满的情形。各国司法界对上述情形的处理意 见尚有分歧，由此暴露出法律适用的复杂状态。在以《纽约时报》为原告起诉 OpenAI 和微软的案件中，原告指责被告未经许可地利用其百万级文章做训练材料，又可输出高度接近原文的片段（甚至整句）。所称的训练过程在此被清楚地划入复制权侵害范围，即把《纽约时报》的文本逐字转移到 OpenAI 的服务器上。德国慕尼黑法院就 GEMA 对 OpenAI 的诉讼已作出较严格立场的判断。根据法院说法，歌词类信息一旦进入模型的参数体系就可再现，用户以简单方式提示模型便能获得歌词，按“记忆”所言属于德国著作权法中“复制”项下的事实，应是非法复制。与之对照的是英国法院对 Getty Images 诉 Stability AI 的宽容态度。法院认为 StableDiffusion 所用的训练方法是参数化、非图像直存的，所生图像不等于既存作品，故不主张把模型训练直接当作侵权行为。

生成式人工智能在版权方面带来的核心问题是可能存在复制或过分借鉴受保护作品的行为，鉴于此，即使先将训练数据的合法性进行搁置，也会造成各种因生成内容与原作太相似，而引发的版权争议。[9]。所以，当用户在给 AI 输入提示词时，如果要生成的是带有明显特点角色，或者是有突出风格的场景内容，那么侵权的风险就会随之增高 [10]。例如，直接让其生成美杜莎、奥特曼等知名角色的图片。在这个过程中，生成结果若包含原作角色的独创性表达，那就对原作品的改编权产生了一定的侵犯。

4 治理新格局

4.1 强化技术治理能力，发展可信 AI

现在技术进步来得如此之快，治理也必须借助技术加以展开，所以技术治理是破局的一种可能方式，目前要着力于建立“以技术对抗技术”的主动防御体系。复旦大学有关人员所提 LmBD 方法对技术治理有启发意义，按此模型把文本与图像作多维结合后，对于修正过的 AI 生成的文本判断准确率可提高 19.68%。若用“数字水印 + 区块链”中双轨溯源的思路处理问题，就可实现隐性识别纳入治理的考虑范围。数字水印属于隐式标识手段，应设计成对裁剪、压缩及编码变换不敏感的鲁棒形式，才适合作为 AI 的生成标记。区块链所依赖的技术可使各阶段都具有信息完整性、不可篡改性，正好构成“内容 DNA”类溯源系统。

此外，要设立“AI 安全沙箱”类的测试环境，把红队对抗训练纳入方法，按受控思路来识别所用模型的各种偏见或伦理缺陷、安全漏洞。对于医疗、司法、金融诸项重要用途的高

ISSN：3105-0085
EISSN：3105-0093

风险 AI 模型来说，必须由第三方做伦理安全审查，待审查合格后再安排实际应用，由此使技术起点上就具有较高可信性与安全性。

4.2 压实平台责任，优化审核机制

在完善治理体系过程中，平台的作用不容忽视，因为平台不仅监管着内容提供者，也连接着用户，通过对其主体责任的落实，才能更好的让治理体系落地。当下，部分电商平台在处理售后纠纷时，为了迎合消费者，片面依赖其提供的静态图片证据，导致商家利益受损。例如，在一些“AI 假图骗退款”的事件中，平台本可以通过商品溯源与物流信息，对消费者信息进行核查，但却没有及时与其他证据交叉验证，这种客观上的纵容，导致了治理环境的进一步恶化。因此，为保证网络社会治理的有效实施，就必须对平台提出要求，督促其升级技术防控与审核体系。

具体来说，平台应针对不同的订单，进行不同的审核。对那些高频申请、异常时段交易的订单，建立分级审核机制。不能让这类申请方仅凭单张图片作为证据，而主动要求其补充提供带时间戳、无剪辑的连续视频证据。最后通过该证据与其他可用信息对照，从而完成交叉对比核验。这样能在技术上增加伪造的难度。

在完善主动审核流程的同时，平台也应提高利用先进技术的能力，把 AI 鉴伪能力作为内在要求，将这种能力整合进自身内容审核流程中。一方面关注技术的动态变化，另一方面接入成熟的 AI 生成内容检测模型，把“发布前检测—疑似内容预警—违规内容拦截”作为一种常态化和自动化的审核流程。通过对用户试图发布的文本、图片、视频等内容进行实时检测，对系统判定为 AI 生成且未按规定添加标识的内容直接拦截，对其他不明内容可标记后交予人工再次处理，完成平台在新时期应尽的责任。至于部分多次发布违规造假内容的账号，平台则应实施梯度化的处罚措施，先禁流禁言，再封号处理，最后纳入跨平台共享的黑名单，从而一步一步引导用户走向正途。

5 迈向生成式治理新范式

生成式人工智能所促成的技术进步本身并不是全部意义的突破，而是网络社会治理逻辑的一次重构 [11]。就目前的治理状况来说，有关政策、有关技术、有关平台已取得一定阶段性的成绩，但多数仍以事后处理为着眼点，部分问题尚待澄清或解决。因此，在未来考虑网络社会治理时应把重心放在事前预防上，借由算法公开及对风险的分析来促使 AI 系统向公共利益靠拢。同时要明白治理不可能一劳永逸地完成，要依现实做渐进调整，努力促成多主体协作的治理状态。唯有如此，技术高速发展的时代才不会失去真相，社会才会有可信、可控的数字文明形态。

参考文献：

[1] 王芳 . 论政府主导下的网络社会治理 [J]. 人民论坛・学术前沿 ,2017,(07):42-53+95.

[2] 徐汉明 , 王玉梅 . 构建独立的网络社会治理法学学科 [N]. 检察日报 , 2019-11-13 (003).

[3] 健全网络生态治理长效机制持续营造风清气正的网络空间 [N]. 人民日报 , 2025-11-30(001).

[4] 杨峰 . 中国共产党领导的互联网治理：历程、成就与经验 [J]. 国家治理 ,2021,(38):37-40.

[5] 郑振宇 . 改革开放以来我国互联网治理的演变历程与基本经验 [J]. 马克思主义研究 ,2019,(01):58-67.

[6] 孙宇 . 网络舆情治理：中国互联网 30 年实践的回顾与展望 [J]. 江西社会科学 ,2024,44(07):188-195.

[7] 赵如涵 , 陈俊旭 . 中美欧网络空间治理比较研究：经验、挑战与未来博弈 [J]. 新媒体与社会 ,1-13.

[8] 闫强 , 蒋婷 , 魏娜 , 易兰丽 . 数智时代下的网络综合治理：研究综述与展望 [J]. 科研管理 ,2025,46(09):46-56.

[9] 吴汉东 . 人工智能生成作品的著作权法之问 [J]. 中外法学 ,2020,32(03):653-673.

[10] 王迁 . 论人工智能生成的内容在著作权法中的定性 [J]. 法律科学 (西北政法大学学报),2017,35(05):148-155.

[11] 张凌寒 , 于琳 . 从传统治理到敏捷治理：生成式人工智能的治理范式革新 [J]. 电子政务 ,2023,(09):2-13.

作者简介：周行虎（2001.02—），男，汉族，河南开封人，硕士研究生，研究方向：新闻传播学。