AIGC 时代网络虚假信息传播特征与协同治理机制研究

陈莹莹

武汉东湖学院，湖北省武汉市，430212；

摘要：AIGC 技术发展为传媒领域带来内容制作便利，却也催生了高仿真、难辨别的批量虚假信息，对信息生态与社会稳定构成威胁。本文以 AIGC 技术发展为切入点研究指出 AIGC 时代网络虚假信息的四大传播特征，主题类型以社会生活类为主，叙事框架采用场景化、细节化与情感化设计，传播主体多元化且隐蔽，传播渠道多平台渗透并以社交平台为核心。AIGC 时代虚假信息大量传播带来了信息环境污染、表达环境失序、虚假信息泛滥等问题。对此，本文提出协同治理机制，构建平台自我监管与国家强制监管结合的治理体系，平台建立成熟的虚假信息辨识处理机制，建立有效的虚假信息标识与救济反馈机制，以确保新时代内容产出的质量。

关键词：新时代；高校思政课；教师队伍建设

DOI：10.64216/3080-1486.25.09.032

2023 年 2 月，中共中央、国务院印发《数字中国建设整体布局规划》指出，建设数字中国是数字时代推进中国式现代化的重要引擎，也是构筑国家竞争新优势的有力支撑。在我国规划建设数字中国这一伟大战略的历史节点，以 AIGC 技术为代表的生成式人工智能技术取得重要突破并被广泛应用到传媒领域，满足了不同用户内容制作和实时互动的需求。然而 AIGC 技术带来机遇的同时也产生了一系列的新问题，AIGC 技术凭借强大的内容生成能力，能够快速批量制作图文、视频、音频等各类虚假信息，且仿真度极高，难以辨别。这些虚假信息借助网络平台的传播优势，迅速渗透到社会各个领域，对信息生态、公共秩序乃至社会稳定都构成了严重威胁。

1 AIGC 时代网络虚假信息呈现新的传播特征

1.1 主题类型——紧跟热点事件，以社会生活类为主

AIGC 生成的虚假信息主题具有明显的高敏感度、高关注度特征，多集中于与公众利益直接相关的社会生活领域，且随社会热点动态变化。从近年案例来看，主要可分为四类：一是公共安全类，涵盖自然灾害（如伪造地震预警）、公共卫生（如虚假疫情信息）、社会治安（如虚假暴力事件），这类主题因直接关联生命安全，易引发恐慌性传播；二是政策民生类，包括教育政策（如伪造高考改革方案）、社会保障（如虚假养老金调整通知）、房产政策（如某地取消限购谣言），这类主题因涉及公众切身利益，传播意愿强；三是名人与公共人物类，如伪造明星、官员、专家的言论或行为，依托名人流量实现快速扩散；四是商业竞争类，多为企业间的恶意攻击，如伪造某品牌产品质量检测不合格报告，通过 AIGC 生成检测数据图表增强可信度。

1.2 叙事框架——场景化、细节化、情感化

AIGC 生成的虚假信息在叙事上高度模仿真实新闻信息的内容逻辑，通过构建场景化、细节化、情感化的框架降低受众警惕性。其一，采用真实媒介叙事模板，虚假新闻会严格遵循 “倒金字塔结构”，开头点明核心事件，中间加入 “受访者”“目击者” 的虚构引言，结尾附上相关部门提醒；虚假科普内容则模仿专家解读框架，引用虚构研究机构名称、假定性数据（如 “某实验室研究表明，某食物致癌率达 90%”）。

其二，强化虚假信息细节填充，例如伪造 “某地食物中毒事件” 时，会具体描述涉事餐厅名称、患者症状（如呕吐、腹泻）、医院接诊人数，甚至加入患者家属的哭诉言论，通过细节增强虚假信息的真实感，让普通受众难以分辨信息真假。

其三，绑定更多受众情感触发点，多采用 “焦虑、愤怒、共情” 三类情感叙事。涉及公共安全主题时强调 “恐慌感”（如再不囤货将断供），涉及社会事件时煽动 “愤怒感”（如某群体被不公平对待），涉及民生主题时引发 “共情”（如贫困家庭因政策变动陷入困境），通过情感驱动提升传播意愿。

1.3 传播主体——多元化、隐蔽化

AIGC 时代网络虚假信息的传播主体呈现多元化、隐蔽化的特征，传播机制则形成 “AI 生成 — 人工分发 — 算法助推” 的闭环。从传播主体来看，可分为三类：一是个人自媒体与流量博主，这类主体通过 AIGC 工具批量生成虚假信息，以 “标题党” 吸引点击，通

过广告分成获利。2024 年，在上海一女童走失事件中，一团伙以“标题党”“震惊体”方式，恶意编造炒作“女孩父亲系继父”“女孩被带往温州”等谣言，该团伙利用 AI 工具等生成谣言内容，通过 114 个账号矩阵，在 6 天内发布 268 篇文章，多篇文章点击量超过 100 万次，造成恶劣社会影响。二是专业水军团队，受雇于企业或特定组织，利用 AIGC 生成统一话术的虚假内容（如批量评论、转发虚假信息），同时使用 AI 生成的 虚拟账号（无真实身份信息）营造多人讨论的虚假氛围。三是无意识传播的普通用户，这类用户因缺乏 AIGC 虚假信息识别能力，误将虚假内容视为真实信息转发至社交圈，成为传播链条中的被动节点。

1.4 传播渠道——多平台渗透、场景化适配

AIGC 生成的虚假信息传播渠道呈现多平台渗透、场景化适配特征，不同类型的虚假信息会匹配不同渠道的传播属性。社交平台（如微信、微博、抖音、小红书）是核心传播阵地：微信朋友圈与微博适合传播长文本虚假新闻、虚假政策解读，依托熟人社交链提升信任度；抖音、快手等短视频平台则是虚假音视的主要载体，如 Deepfake 生成的虚假人物视频、AIGC 制作的虚假事件短视频，通过“短平快”的形式快速触达用户；小红书、知乎等平台则更易传播虚假科普、虚假体验分享（如 AI 生成的“某产品使用测评”），利用平台知识分享、生活记录的定位降低受众防备。

此外，“私域渠道”与“垂直社群”成为传播补充。部分虚假信息会通过微信群、QQ 群等私域渠道传播，依托熟人信任实现精准扩散（如虚假“学区房政策”在家长群传播）；垂直社群（如养生群、股民群）则会定向传播相关主题虚假信息（如“AI 生成的股市内幕消息”在股民群扩散）。值得注意的是，部分 AIGC 虚假信息还会通过跨平台流转扩大范围，例如先在抖音发布虚假短视频，再将视频截图与文字解读同步至微博、微信，形成多渠道联动传播。

2 AIGC 时代网络虚假信息传播危害

2.1 自主生成虚假内容，破坏网络生态

AIGC 技术的核心优势在于能够基于算法和数据自主生成内容，这一特性也为虚假信息的批量生产提供了便利。在缺乏有效监管和约束的情况下，部分主体利用 AIGC 工具，无需专业的内容创作能力，就能快速生成大量虚假新闻、谣言、误导性评论等内容。这些自主生成的虚假内容往往具有较强的迷惑性，其语言风格、逻辑结构与真实信息高度相似，普通用户难以通过常规方式辨别。

从传播规模来看，AIGC 生成的虚假内容能够在短时间内实现海量传播。例如，在重大公共事件发生后，一些别有用心的人会利用 AIGC 技术编造虚假的事件进展、伤亡数据等信息，这些信息借助社交媒体平台的裂变式传播机制，迅速扩散至全网。大量虚假信息的泛滥，挤占了真实信息的传播空间，导致用户陷入“信息迷宫”，难以获取准确、有效的信息，严重破坏了网络信息生态的健康与有序。

2.2 人为虚假信息引导，网络表达环境失序

在 AIGC 时代，网络虚假信息的传播不再仅仅是单纯的内容生成与扩散，还伴随着人为的刻意引导，进一步加剧了网络表达环境的失序。部分个人或组织为了实现自身的利益诉求，如吸引流量、操控舆论、打击竞争对手等，会利用 AIGC 技术制作针对性的虚假信息，并通过精心策划的传播策略，引导公众的认知和舆论走向。

一方面，这些主体会通过 AIGC 技术生成带有强烈情感倾向和误导性的内容，如虚假的负面爆料、极端的观点论述等，刺激公众的情绪，引发非理性的讨论和争议。另一方面，他们还会利用 AIGC 生成的“水军”账号，在网络平台上批量发布评论、点赞、转发等，营造出虚假的舆论氛围，干扰正常的舆论判断。这种人为引导的虚假信息传播，不仅会导致公众对事件的认知出现偏差，还会引发网络暴力、群体对立等问题，破坏了公平、理性、有序的网络表达环境，严重影响了网络空间的和谐稳定。

3 AIGC 时代网络虚假信息协同治理机制构建

针对 AIGC 时代网络虚假信息传播的特征与危害，单一的治理主体和治理手段难以有效应对，需要构建政府、平台、社会公众等多方参与的协同治理机制，形成治理合力，共同遏制虚假信息的传播。

3.1 建立平台自我监管与国家强制监管相结合的治理体系

平台作为网络信息传播的重要载体，应承担起自我监管的主体责任。一方面，平台要加强对 AIGC 技术应用的管理，建立严格的 AIGC 内容生成审核机制，对用户使用 AIGC 工具生成的内容进行前置审核，从源头上减少虚假信息的产生。例如，平台可以开发专门的 AIGC 内容检测技术，对生成内容的真实性、合法性进行自动筛查，对疑似虚假信息的内容进行人工复核，确保只有符合规范的内容才能在平台上传播。

另一方面，国家应加强对网络平台的强制监管，完善相关法律法规体系。制定针对 AIGC 时代网络虚假信息传播的专门法律，明确虚假信息的界定标准、传播主

体的法律责任以及治理措施等，为虚假信息治理提供明确的法律依据。同时，监管部门要加大对平台的监督检查力度，对未履行自我监管责任、导致虚假信息大量传播的平台，依法予以处罚，如罚款、责令整改、暂停服务等，倒逼平台落实监管责任。通过平台自我监管与国家强制监管的有机结合，形成 “内外共治” 的治理格局，有效提升虚假信息治理的效率和效果。

3.2 平台建立成熟的内部虚假信息辨识与处理机制

平台内部虚假信息辨识与处理机制的成熟与否，直接关系到虚假信息治理的及时性和有效性。首先，平台要加大技术研发投入，提升 AIGC 虚假信息的辨识能力。利用人工智能、大数据、区块链等先进技术，构建多维度的虚假信息辨识模型。例如，通过分析 AIGC 生成内容的语言特征（如异常高频词汇、逻辑矛盾点）、数据来源（如虚构引用源）、逻辑结构（如不符合真实事件发展规律）等，识别出其中的虚假信息。

其次，平台要建立高效的虚假信息处理流程。一旦发现虚假信息，要立即采取删除、屏蔽、降权等措施，阻止其进一步传播。同时，要及时向用户发布虚假信息预警，告知用户相关信息的虚假性，引导用户正确认知。此外，平台还应建立虚假信息处理反馈机制，及时向用户反馈虚假信息的处理结果，增强用户对平台治理工作的信任和支持。通过建立成熟的内部虚假信息辨识与处理机制，平台能够快速、有效地应对虚假信息传播问题，减少虚假信息带来的危害。

3.3 建立有效的虚假信息标识与救济反馈机制

虚假信息的标识与救济反馈，是治理链条的 “最后一公里”，其核心目标是通过清晰的标识降低用户误判风险，通过完善的救济机制弥补虚假信息造成的损害，形成 “识别-警示-补救”的闭环治理。

在虚假信息标识机制构建上，需遵循 “清晰化、差异化、场景化”原则，确保用户能快速识别虚假信息。首先，应制定统一的 AIGC 虚假信息标识标准，明确不同类型虚假信息的标识形式。对文本类虚假信息，在内容标题前添加红色 “虚假信息”标签，并在正文开头标注“本内容经核查为 AI 生成虚假信息，具体核查依据见链接”。其次，需根据虚假信息的危害程度实行差异化标识，对 “低危害型”（如娱乐八卦类虚假图文），采用基础警示标识；对 “中高危害型”（如涉及公共安全、民生政策的虚假信息），在标识基础上，还需在用户打开内容时弹出风险提示弹窗，明确告知 “该信息可能影响您的决策，建议参考官方渠道信息”，强化警示效果。

在虚假信息救济机制设计上，需聚焦用户权益保障与损害弥补，覆盖不同主体的救济需求。针对普通用户，需建立认知纠偏与权益受损救济双轨机制，一方面，对曾接触过虚假信息的用户，平台可通过个性化推送向其精准推送权威辟谣内容；另一方面，若用户因轻信 AI 虚假信息遭受直接损失（如因虚假投资推荐被骗、因虚假商品宣传购买劣质产品），可通过平台维权通道提交损失证明（如转账记录、购物凭证），平台需协助用户联系涉事账号所属主体，协商赔偿事宜；对协商无果的，可联动司法部门提供法律援助指，降低用户维权成本。

参考文献

[1]唐芳芳.AIGC 虚假信息的内容特征与治理路径——基于对 79 例 AIGC 虚假信息的内容分析[J].新闻前哨,2025,(12):50-51.

[2]夏颖,赵子娇.AIGC 模式下假新闻生产进路探析及反思[J].传媒论坛,2025,(10):24-26.

[3]王天铮,袁韬.AIGC 时代网络空间虚假信息的元规制研究[J].当代传播,2025,(02):35-42.

[4]李瑾颉,吴联仁,齐佳音.对抗AIGC虚假信息的行为助推与认知干预研究[J].青年记者,2025,(05):53-57.

[5]邱均平,张廷勇,徐中阳.数智时代下UGC与AIGC虚假健康信息特征的对比研究[J].情报理论与实践,2025,48(05):1-12.

[6]张新生,王润周,马玉龙.AIGC 背景下虚假信息治理挑战、机会与策略研究[J].情报科学,2025,43(05):78-88+96.

[7]刘高含.生成式AI虚假信息风险类型化分析及治理路径[J].科技传播,2024,16(20):109-113.

[8]漆晨航.生成式人工智能的虚假信息风险特征及其治理路径[J].情报理论与实践,2024,47(03):112-120.

作者简介：陈莹莹（1994），女，汉族，山东省淄博市，助教，硕士研究生，网络与新媒体，武汉东湖学院。

基金项目：2025 年武汉东湖学院青年基金项目“AIGC 时代网络虚假信息传播特征与协同治理机制研究”（2025dhsk041）.