

颠覆性技术下涉华舆论特征、风险及应对

杨亚^{1, 2}

(1. 桂林电子科技大学, 广西 桂林 541000; 2. 广西师范大学, 广西 桂林 541000)

摘 要: 随着人工智能、深度伪造、算法推荐等颠覆性技术的快速迭代, 全球舆论生态正在经历深刻重塑。在涉华国际舆论场中, 颠覆性技术一方面为讲好中国故事提供了新工具、新渠道, 另一方面也催生了前所未有的舆情风险。虚假信息生成与传播的工业化、算法偏见导致的认知锁定、深度伪造技术引发的真相危机、社交机器人操纵舆论的虚假共识, 构成了当前涉华舆情的主要风险, 应该从技术治理、制度完善、内容创新、国际合作等多维度进行系统性应对。

关键词: 颠覆性技术; 涉华舆论; 深度伪造; 人工智能

DOI: <https://doi.org/10.65196/b6prwg52>

“颠覆性技术源自颠覆性创新, 具有创新性、突破性和替代性等特征”¹, 如今, 颠覆性技术的更迭速度加快, 对国际舆论场带来重大的影响和冲击, 使其呈现出高度复杂化和动态化特征, 以往主要依靠议程设置和框架理论的舆论引导模式, 当下正遭遇全新挑战。关于涉华舆论, 一直以来都伴随着意识形态偏见和地缘政治博弈, 在技术加速变革的影响下, 也呈现出一些新的特点和趋势。在此背景下, 探索涉华舆论新的特征和潜在风险, 积极研究应对之策, 具有重要的理论和现实意义。

一、颠覆性技术下涉华舆论的特点

“西方国家主导的对华舆论攻势呈现绵延不断的发展态势”², 颠覆性技术的介入, 使涉华舆论的生成逻辑和传播规律发生了重大变化。相比于传统媒体时代自上而下的舆论生成模式, 当下的涉华舆论呈现出更加复杂的特征。

(一) 舆论内容生产更趋自动化

在智能化时代之前, 涉华舆论的生成, 更多依靠的是媒体记者、意见领袖, 或者是普通网民的创作, 以人工主导为主, 这种情况下由于受到时间和精力牵制, 信息的产生相对有限。但是, 随着生成式人工智能等颠覆性技术的深度发展, 这一生成逻辑和限制被彻底打破, 信息的生成实现了爆发式增长。生成式人工智能只需要特定的指令提示, 就能够在极短时间内生成海量涉华内容, 可以涵盖文本、图像、音频、视频等多重形式。由于生产不再受人力制约, 传统水军也更新为机器水军。随着技术的升级发展, 如今只需设定特定的内容框架, 大模型就能够生成各种视角和不同风格的叙事文本, 实现规模化和批量化的信息制造。

(二) 信息传播呈现裂变式扩散

颠覆性技术相比传统技术的跨越式发展, 大大降低了信息的生产以及分发成本, 因而在涉华舆情的传播上, 也由线性增长变成了裂变式扩散。在社交平台和即时通讯工具上, 由于算法推荐的融入, 使得一条信息可以在非常短的时间内就传达到全球各地, 覆盖到全世界用户。在 Twitter、

基金项目: 广西高校中青年教师科研基础能力提升项目“颠覆性技术下涉华舆情风险及引导研究”(项目编号: 2024KY0188); 广西高校大学生思想政治教育理论与实践研究课题“数智时代主流意识形态场景化叙事研究”(项目编号: 2025SZ009); 教育部人文社会科学研究青年基金项目“数字媒介时代主流意识形态具象传播及认同研究”(项目编号: 25YJC710076); 广西哲学社会科学研究项目“数智时代铸牢中华民族共同体意识具象传播研究”(项目编号: 24KSF001); 广西研究生教育创新计划项目“媒介技术演进下主流意识形态叙事的场景适配性研究”(项目编号: YCBZ2026108)。

作者简介: 杨亚(1988-), 男, 博士研究生, 助理研究员, 研究方向为网络思想政治教育。

¹ 林雪莹, 陈志勇. 颠覆性信息技术视域下意识形态传播的风险转向[J]. 东南学术, 2023, (3): 217-224.

² 徐明华, 晏栗群, 魏子瑶. 隐蔽式并发与选择性煽动: 西方对华计算宣传的生成逻辑与运作机制[J]. 当代传播, 2024, (3): 55-60.

Facebook、TikTok 等主流平台上，推荐算法往往根据用户的网络行为，实时调整推送的策略，当涉华内容被标记为高热，就会获得更大流量，形成“滚雪球”效应，获得巨大的传播势能，实现信息传播的指数级扩散。

（三）信息推送更具针对性精准性

在大数据和机器学习技术的推动下，用户画像技术已经被深度应用到信息传播领域，因此，作为舆论操控者，能够实现对特定群体的精准打击。通过对用户的地理位置、语言偏好、浏览历史、社交网络、情感倾向等数多个维度的分析，算法可以识别出所谓的目标人群，进而对那些本来就拥有偏见的人群，或者认知基础较为薄弱的对象开展定向信息输送。操控者基于受众画像，对各类不同人群推送符合他们认知的内容，这种千人千面的信息投喂策略，极大地提高了舆论诱导的成功率，使负面信息的渗透深度和转化效率远超传统宣传手段。

（四）传播方式更趋隐蔽性

人工智能尤其是生成式人工智能的发展，深度改变了受众获取涉华信息的方式，他们不再只是通过传统媒体获取，而是基于大语言模型的对话功能，越来越倾向于直接向 AI 提问。众多被曝光的案例显示，国外大模型在回答涉华相关问题，尤其是敏感问题时，其观点答案的立足点并不是客观中立态度，而是基于其受训的语料库以及对齐算法的规则。对话式 AI “在满足用户信息需求的同时，以一种隐蔽的方式将信息中的意识形态向对象国渗透，从而自然、有效、深入地影响用户，规避国际矛盾和冲突”³，让人在不知不觉接收其带有意识形态偏见的输出。由于这种方式更具隐蔽性，受众接收时往往没有防备，所以通常比传统媒体的宣传更具说服和渗透效果。

二、颠覆性技术下涉华舆论的风险

颠覆性技术不断演化形成了新的舆论特点，也给涉华舆情带来了新的挑战，虚假信息生成的工业化规模化、算法偏见导致的叙事偏差、深度伪造带来的真相危机以及社交机器人的舆情操控等，这些风险的相互叠加，共同作用，构成了颠覆性技术下涉华舆论面临的主要挑战。

（一）虚假信息生成规模化扰乱舆论秩序

生成式人工智能的突破性进展，为虚假涉华信息的产生提供了便利，以往需要花费大成本的虚假信息生产，如今变得较为容易。比如对于一些敏感涉华议题，操作者仅仅需要在相应平台输入指令，就能够在短时间内生成详细的看似真实的“事件经过”，甚至连图片和视频，也可以在数秒内就生成所谓的“现场画面”，生成的这些内容，不仅仅成本低，而且通常较难通过传统的反向搜索辨别真伪，因为 AI 生成的内容，往往来自所谓的全新创造。而更具有挑战性的是，AIGC 支持大规模的定制，以及变体生成，同样的虚假叙事框架，往往能够较为轻易的生成多个不同的版本，可以投放传播到具有不同语言以及不同文化背景的受众当中。“数字虚假信息的产生和运作是一场巨大的‘公共传播危机’”⁴，显著增加了事实核查的难度，破坏了良好的网络生态和舆论秩序，给舆情防治带来挑战。

（二）算法偏见嵌入导致的叙事偏差

算法推荐已经融入信息传播的方方面面，表面上看，算法以用户兴趣和相关性为优化目标，但实际上，看似中立的算法内嵌着多重偏见。比如在数据训练方面，西方的主流推荐所使用的训练数据，主要来源于西方的英文互联网，其中有关中国的语料，被诸多西方叙事框架所主导，许多带有意识形态偏见，容易将中国与一些负面标签形成高度语义关联，所以当用户搜索相关信息时，算法往往倾向于推荐带有偏见的匹配内容。另外，大量研究也表明，负面信息在社交媒体中往往能够获得更高的关注度和点击率，算法的核心指标天然倾向于放大负面内容，因而对于涉华议题，意味着负面消息，不论其真假，往往会获得算法推荐的更高权重。而且通常情况下，算法会根据用户的过往行为，进行个人所喜好内容的强化推荐，用户不知不觉陷入特定的信息循环之中，使得涉华舆情的极化程度加剧，从而导致中间地带和理论讨论空间的压缩。

³ 吴璜, 孙鸣伟. AIGC 时代涉华国际舆论的演变、风险与敏捷治理——以 ChatGPT 为例[J]. 福建师范大学学报(哲学社会科学版), 2024, (5): 105-115+171-172.

⁴ 张梦, 罗昕. 西方涉华计算宣传的动向、风险与应对[J]. 统一战线学研究, 2023, 7 (2): 115-124.

（三）深度伪造技术造成真相危机

“眼见为实”“有图有真相”，这是传统认知中公众判断真假信息的基本准则，但是深度伪造技术的普及，却正在冲击这样的底层认知。深度伪造通过其深度学习模型，能够生成高度逼真的合成内容，可以实现人的面部表情、声音口型，以及身体动作的移植。具体到涉华舆情的风险方面，一是对国家领导人和政府官员的虚假攻击。比如利用一些公开的视频讲话素材，就能够轻易生成抹黑中国形象的伪造视频，这些视频一旦进入传播，即便事后被证伪，仍然会造成难以挽回的声誉损害和信任危机。二是对敏感事件的情境重构。比如对于一些西方势力热衷炒作的边疆问题，深度伪造技术可以凭空生成一些根本不存在的现场冲突对抗视频，而这些具有煽动性的视频往往传播速度很快，给官方辟谣带来挑战。需要警惕的是，如若深度伪造视频泛滥，公众可能会陷入一种普遍怀疑的困惑状态，在模糊认知的影响下造成真相危机，影响国家和社会的安全稳定。

（四）社交机器人制造虚假共识操纵舆论

社交机器人是由算法控制的自动社交媒体账号，能够模拟人类用户的行为，包括发布内容、点赞、转发、评论、关注等。在涉华舆情的形成过程中，社交机器人扮演着“放大器”和“造势者”的双重角色。社交机器人可以被大规模部署来扩散特定的涉华负面信息。比如一条涉华负面帖子，如果获得数千个机器人账号的转发和点赞，就可能被平台的算法判定为“热门内容”而推送给更多真实用户，形成虚假热度的错觉。以往的相关报道和研究显示，在多个涉华热点事件中，社交机器人的活跃度显著高于同期其他议题。更值得注意的是，这些机器人账号的行为模式呈现出高度协调的特征，他们集中在特定时间窗口内爆发式活动，使用高度一致的话语框架，并且能够自适应地回应真实用户的质疑，这说明其背后有着明确的组织方和技术支持，而非零散的个体行为。

三、颠覆性技术下涉华舆论的应对

面对颠覆性技术带来的系统性舆情风险，需要对传统的引导应对方式进行全面升级，构建一套涵盖技术、制度、内容、国际协作等完善的多层次引导体系。

（一）加大技术研发力度，以技术对抗技术赢得舆论战

面对颠覆性技术驱动的舆情风险，应加快提升技术层面的主动防御能力，要推进国产 AI 大模型研发，着力培育具备国家正义属性的 AI 系统，以识别与对抗散布反华信息及二次传谣的恶意为行。⁵针对深度伪造技术的风险，应进一步提高检测技术，开发具有更高识别能力的检测算法，同时，要建立起覆盖主要视频平台的深度伪造实时监测系统，在虚假内容大规模传播之前发出预警。同时，研究提升 AIGC 内容溯源与标记技术，推动生成式 AI 服务提供商的透明化责任，要求其在生成内容中嵌入不可见水印或数字签名，以便检测工具能够快速精准识别。另外，还要开发能够实时识别机器人账号集群的算法，分析其发帖频率、互动对象的时间分布、内容重复率等行为模式，并与社交媒体平台建立协同处置机制，当检测到针对涉华议题的大规模机器人操纵行为时，应立即进行限流、打标签或批量封禁。

（二）建立完善法律法规和制度体系，积极参与国际标准制定

除了加强技术手段之外，还应该在制度手段上下功夫，通过建立明确的法律责任和行业标准来强化对有关技术的监管。在立法层面，应加快出台完善针对深度伪造和 AIGC 的专项法规。要求任何利用深度伪造技术生成的涉政类内容，必须在显著位置标明“此为合成内容”，并披露制作方信息。对于故意制作、传播深度伪造涉华虚假信息的行为，应明确行政和刑事责任，并为跨境追责和司法协作提供法律依据。在标准层面，应推动制定国际通行的 AI 生成内容标识标准，积极参与相关国际机构的规则制定，提出中国方案，争取将内容溯源和标记要求纳入国际标准。而在应急机制方面，应建立涉华重大舆情风险的跨部门快速响应机制，当检测到针对中国的深度伪造视频或 AIGC 虚假报道大规模传播时，应该启动应急响应方案开展应急核查和紧急辟谣，进行及时

⁵ 陆建平, 党自强. AIGC 时代虚假信息的制造传播与国家安全及公民权益的维护[J]. 浙江大学学报(人文社会科学版), 2024, 54(5): 42-58.

纠偏。

（三）注重内容和渠道的创新，提升在国际社会的话语主动权

在信息过载和注意力碎片化的时代，面对涉华舆情的新情况新特点，还必须要加强注意力和话语权的争夺，提供更有吸引力、感染力、易于传播的信息产品。一方面，中国对外传播内容需从“宏大叙事”转向“微观叙事”，要通过具象化和生动性的内容，符合国际受众的接受习惯，让他们穿透信息茧房束缚，形成对中国的客观认知。另一方面，传播形式应适配颠覆性技术时代的媒介特征。通常情况下，短视频的时长较短，信息密度极高，前几秒就能抓住观众注意力。传统的新闻稿件式风格在一些视频类传播平台上适配度不高，因此要组建专业团队研究平台算法偏好，制作节奏明快、视觉冲击力强、情绪传达准确的多模态内容。此外，还应主动使用 AIGC、虚拟现实、增强现实等技术，制作高质量的正面内容，通过增强沉浸感和体验感，为其展现真实的情况，建立更加广泛共识，增强情感共鸣。

（四）加强国际交流与合作，构建公正合理的治理体系

颠覆性技术驱动的涉华舆情风险具有明显的跨境特征，虚假信息的制造者服务器、传播渠道和攻击目标可能分别是不同国家，影响的受众遍布全球，基于这样的现实，需要加强国际合作，共同应对复杂挑战。中国应积极与主要国家建立打击虚假信息的对话机制和执法合作渠道，在打击深度伪造、社交机器人操纵等明确恶意行为上，寻找各方存在共同利益，通过加强技术合作逐步建立互信。同时，要在重大国际合作组织和平台中，倡导建立全球性的事实核查网络和深度伪造内容数据库，通过这一数据库，可以记录已被证实为虚假的相关内容，供全球各平台和事实核查机构查询比对。此外，还应在全球数字治理中提出兼顾安全、发展与人权的治理理念，与广大发展中国家形成话语联盟。针对西方主导的叙事中存在的风险和问题，中国可以倡导“负责任的人工智能”“技术向善”“公平的信息秩序”等替代性叙事，争取更多国家的认同和支持。

Characteristics, Risks, and Responses of Public Opinion Related to China Under Disruptive Technologies

YANG Ya^{1,2}

(1. Guilin University of Electronic Technology, Guilin, Guangxi 541000, China; 2. Guangxi Normal University, Guilin, Guangxi 541000, China)

Abstract: With the rapid iteration of disruptive technologies such as artificial intelligence, deepfakes, and algorithmic recommendations, the global public opinion ecosystem is undergoing profound reshaping. In the international public opinion sphere involving China, disruptive technologies on the one hand provide new tools and channels for telling China's story well, and on the other hand, they have given rise to unprecedented public opinion risks. The industrialization of false information generation and dissemination, cognitive lock-in caused by algorithmic bias, the truth crisis triggered by deepfake technology, and the false consensus manipulated by social bots constitute the main risks of current public opinion involving China. Systematic responses should be taken from multiple dimensions, including technological governance, institutional improvement, content innovation, and international cooperation.

Keywords: disruptive technology; China-related public opinion; deepfakes; Artificial Intelligence