

AI 谣言传播之组态路径研究

——基于 35 个案例的定性比较分析^①

向安玲¹, 张诗瑶², 程栋³

(1. 中央民族大学新闻与传播学院, 北京 100081;

2. 清华大学新闻与传播学院, 北京 100084;

3. 西京学院传媒学院, 陕西西安 710062)

摘要: 生成式 AI 驱动下, 谣言传播已从“异态”转变为“常态”, 从“单模态”演变为“多模态、跨模态”, 从“合成事实”进化到“伪造舆论”。探讨 AI 谣言的传播机制和影响因素, 对于 AI 谣言识别和治理具备支撑作用。针对近年来传播热度较高的 35 个 AI 谣言传播案例, 使用定性比较分析 (QCA) 对其传播热度的多层要素进行挖掘。研究发现, 目前 AI 谣言在大众化传播上存在三级断层, 前 20% 的高热度谣言占据了 84.7% 的大众流量; 从谣言传播影响因素来看, 内容特性 > 媒介模态 > 受众参与 > 信源构成; 从组态路径来看, 草根创作的“强情感”叙事和权威背书的“强冲突”框架构成了 AI 谣言大众化传播的两种典型模式; 在 AI 谣言识别和治理过程中, 多模情绪交织与大众二次创作构成了破局痛点与难点。

关键词: AI 谣言; 传播机制; 定性比较分析; 演化博弈理论; 组态路径

中图分类号: G206

文献标识码: A

文章编号: 2096-8418 (2025) 01-0046-11

随着 Deepseek、ChatGPT 等媒介技术的发展, 人工智能时代的谣言已从“异态”转变为“常态”, 从“单模态”演变为“多模态、跨模态”, 从“合成事实”进化到“伪造舆论”。AI 技术已经显著改变信息传播的景观, 不仅增强了谣言信息的迷惑性, 还在认识论和本体论上造成信息割裂, 引发了深刻的媒介生态学变革, 强化了信息的“无界性”和“多维性”。AI 生成谣言内容, 不再是单一信息的复制或再现, 而伴随着多重信息的选择、重组与再创造, 使得谣言内容具有更高的逼真度和更复杂的情感色彩。^[1] 此外, AI 技术在处理大规模数据时的高效能和适应性, 使得个性化谣言信息的生成和跨圈传播成为可能, 从而在微观层面重新塑造了公众的认知结构和社会

互动方式。这种技术驱动的信息流变动, 不仅挑战了传统的信息验证机制, 也对社会公共信任的构建提出了新的要求。^[2] 在此背景下, 探究 AI 在信息环境中如何同时扮演构建性与破坏性角色, 成为亟待解答的关键问题。

针对 AI 谣言的扩散及其对社会秩序的潜在破坏, 各国政府已推出系列防范措施与治理政策。我国相继推出《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》等政策规定, 并持续性开展“清朗·打击网络谣言和虚假信息”专项行动。尽管 AI 谣言的识别与治理已引起各界重视, 但当前 AI 谣言传播的影响因子尚缺乏系统性梳理, AI 谣言的传播逻辑也有待深入研究。尤其是在全球化社交媒体环境下, AI 谣

基金项目: 国家自然科学基金青年项目“面向人工智能生成内容的风险识别与治理策略研究”(72304290); 中国博士后科学基金面上资助项目“基于机器博弈的 AI 谣言识别与治理研究”(2024M751595); 全军军事理论科研计划重点项目“AI 深度伪造新闻传播机制与认知影响研究”(24GDJ30068B)。

作者简介: 向安玲, 女, 讲师, 博士; 张诗瑶, 女, 博士后; 程栋, 男, 教授。

① 此文系第五届西京智媒体国际论坛获奖优秀论文。

言的跨境流动和大众化再造使其识别和治理难度升级。AI 谣言为何得以大众化传播? 有哪些因素促成了 AI 谣言的“跨圈”传播? 是否存在谣言传播的关键性“阻断因素”? 治理 AI 谣言应该从哪些层面切入? 以上问题有待进一步探讨。

在此背景下, 本文选取近年来 AI 谣言传播典型案例, 采用定性比较分析挖掘影响 AI 谣言传播的多层要素。一方面, 针对 AI 谣言传播的影响因子进行梳理, 尝试搭建 AI 谣言传播机制的系统性框架; 另一方面, 从“信源组织性”“谣言主体知名度”“内容情绪调性”“内容需求度”“媒介模态复杂度”与“受众参与度”6 个维度出发, 针对 AI 谣言传播的多层次因素进行定性比较分析 (QCA), 由此提炼影响 AI 谣言传播的组态路径, 并对其背后的传播逻辑与运作机制进行探讨, 以期提高社会对 AI 谣言的整体认知度和治理能力。

一、文献回顾

(一) AI 谣言概念界定与传播特性

谣言作为一个学术概念, 已有超过半个世纪的研究历史。由于谣言是一种复杂的社会现象, 试图为其提供一个绝对性的定义仍具挑战性。有观点认为谣言是主观恶意捏造的不真实信息,^[3] 是缺乏事实根据的虚假消息,^[4] 甚至认为凡是谣言一定是虚假的、非事实的, 如果是真实的就不能称之为谣言。^[5] 然而, 部分学者开始注意到谣言在社会中的正面功能。例如, 将其视为世界上最古老的传媒, 并将它看作是一种反权力,^[6] 将其理解作为一种对抗权力的社会表达。^[7] 与此同时, 有学者指出谣言的虚构与真相只是一个结果判断, 这可能忽视了谣言传播过程中的关键机制。^[8] 可见, 国内外学界对于谣言的定义尚无统一共识。

随着人工智能技术的发展, AI 谣言成为一个新的研究热点。AI 谣言具有幻觉效应和快速迭代的特点, 并通过机器学习、深度伪造、社交机器人等技术实现多模态融合传播。这些谣言通过夸张、对立和引人入胜的内容操控语言, 以激发用户的情感反应, 从而促使其分享和讨论。^[9] AI 谣言的自动化特性使其能够在社交媒体平台上迅速

传播, 并精准定位特定受众。^[10] 算法技术下的谣言信息“群体驱动”和“信息茧房”效应使谣言传播范围更广、影响更深远。^[11] 此外, AI 谣言还对社会、政治和商业构成了重大威胁, 如干扰选举、破坏公众对官方信息的信任、引发网络安全等问题等。^[12] 这些特性使 AI 谣言从单一传播主体转变为多主体协同的谣言传播, 并推动了从中心化传播到分布式发布的演变。因此, 研究 AI 谣言的传播机制, 尤其是在人工智能技术背景下, 成为维护国家信息安全的重要课题。

基于上述分析, 本文将 AI 谣言定义为利用人工智能技术生成、传播或增强的, 旨在影响公众认知、情绪和行为的未经证实的虚假信息。这一定义突破了传统谣言的认知范畴, 强调了 AI 技术赋予谣言的新特征、新危害和新挑战。

(二) AI 谣言的传播影响因素

谣言传播的影响因素在传播学领域有着丰富的理论基础。一个经典的公式是“ $R=I \times A$ ”, 它指出谣言的产生与传播依赖于信息的匮乏、不安与忧虑以及危机情境等关键条件。这一公式揭示了谣言传播的内在动力: 当任何一个因素为零时, 谣言便难以产生和传播。^[13] 同样, 另一个公式“谣言=事件的重要性 \times 事件的模糊性”也强调了事件的重要性和不确定性在谣言传播中的作用。

从内容层面来看, 谣言的传播力与其内容的分布、叙事结构、议题构建以及标题特征密切相关。通过巧妙运用故事元素, 结合情感和理性的表达, 以及多模态素材的混合使用, 谣言在传播中更具吸引力。^[14] 此外, 叙事策略也会影响谣言传播的热度。研究发现, 86% 的谣言标题采用非人称叙事, 以避免出现明确的人称, 从而增强客观性。^[15] 谣言主题、指涉目标、佐证证据、内容主张 4 个方面在谣言的传播过程中起着重要作用。特别是在突发性重大公共卫生事件中, 有研究发现, 谣言主题多集中在病例与防疫措施上, 指涉目标的时间、地点和对对象较为明确, 佐证证据多突出权威性与图像证据, 而内容主张则以分享新知为主, 行动建议为辅。^[16] 社交媒体上的健康谣言表现出鲜明的主题特征、情感特征、形式特征

及用户特征。健康谣言的主题贴近生活,情感表达更为强烈,且形式愈发多样化。^[17]不同学者的研究一致表明,谣言的主题特征与情感特征是影响其传播效果的重要因素。在传播媒介层面,当用于虚假信息时,视觉效果可能成为虚假信息的强大传播者,因为它们通常被认为比单独的文本线索更可信、更准确。多模式虚假信息被认为比文本虚假信息更具可信度,但仅在虚假信息涉及攻击难民参与恐怖主义时,这种差异才显著。^[18]关于媒介模态复杂度对谣言传播的影响,学界仍存在争议。在谣言传播过程中,参与者的知识储备、声誉及影响力同样会影响其行为选择,进而决定谣言的传播路径和广度。^[19]

上述研究为理解 AI 谣言的传播机制奠定了坚实的理论基础,并揭示了多种关键因素。基于这些理论成果,本文选择了以下 6 个变量作为分析 AI 谣言传播的条件变量:信源组织性(X1)、谣言主体知名度(X2)、内容情绪调性(X3)、内容需求度(X4)、媒介模态复杂度(X5)以及受众参与度(X6)。这些变量将为进一步研究 AI 谣言的传播路径提供重要的分析框架。

(三) AI 谣言的社会影响与破局关键

当下,AI 技术的应用改变了信息生成模式,谣言变得“定制化”“精准化”“智能化”。^[20]快速生成的谣言信息在越来越拟人化的技术背景下,使得传统平台难以连续识别虚假信息,^[21]对整体的 AI 虚假信息治理体系提出了挑战。^[22]AI 谣言不仅侵蚀公众对信息源和机构的信任,导致社会信任度下降,还可能引发社会恐慌,扰乱公共秩序,甚至导致集体行为的极端反应。此外,还可能造成经济损失,影响市场稳定和商业信心,对政府和公共机构的公信力构成威胁,削弱其在公众中的威信和管理效能。^[23]除此之外,AI 谣言还能够煽动恐怖和暴力活动、挑拨国家与社会矛盾,^[24]甚至扰乱公共话语,^[25]控制战争的叙事。^[26]

目前,关于 AI 谣言的研究大多集中于计算机科学领域的生成式、技术性、算法类研究,而社会科学领域的理论性和管理类研究则相对匮乏。在研究视角方面,AI 谣言的研究主要涉及 AI 谣

言的检测与识别以及数学建模和仿真。相较之下,社会科学在 AI 谣言领域的探讨多采用思辨性质化研究,缺乏实证量化的支撑。目前,对于 AI 谣言的识别研究主要限于对潜在的错误信息进行标记(分类任务),未能充分解决问题。

综上,既有研究侧重于技术性解决方案,往往忽视谣言传播过程中的社会行为和心理因素,缺乏整合性视角。AI 谣言研究的局限性表明,有必要整合计算机科学、社会学、心理学等多学科知识,以更有效应对 AI 谣言的复杂性和隐蔽性。在此背景下,本研究从更具象的颗粒度分析 AI 谣言的传播因素及其组态路径,探讨如何有效识别和治理 AI 谣言。具体研究问题如下:

研究问题一:目前 AI 谣言的“信源组织性”“谣言主体知名度”“内容情绪调性”“内容需求度”“媒介模态复杂度”与“受众参与程度”在传播影响上是否存在显著差异?

研究问题二:是否存在某种属性可构成 AI 谣言大众化传播的必要条件?如果存在,具体包括哪些?

研究问题三:是否存在某些组合条件(组态)可影响 AI 谣言的传播热度?AI 谣言大众化传播的最常见组态是什么?

二、研究方法

(一) 数据获取

为了避免平台既有用户规模、生态特性、用户圈层对研究结果造成的干扰,本文针对同一平台上传播的 AI 谣言进行分析。目前市场上热度相对较高的 AI 谣言主要活跃在网页端、微信公众号和微博平台,其中微博平台集成了图文、视频等多模态交互方式,且除了单纯的信息交互和内容输出,微博平台的话题使得 AI 谣言的传播影响力得以凸显。

在具体 AI 谣言样本选择上,一方面,通过关键词检索,对微博平台上带有“AI 谣言”“AI 假新闻”“AI 诈骗”“AI 虚假信息”等标签的话题进行样本初筛,并结合文献综述中对于 AI 谣言的概念界定和核心社会线索,根据人工判断进行样本细筛;另一方面,根据网络公开报道数据对当下传播热度较高的 AI 谣言案例进行补充。为了尽

量规避样本账号“刷量”等行为的干扰，一方面对样本话题讨论中的粉丝规模、发布活跃度、交互频率等多类数据进行多维统计，降低了单一指标水分对研究结论带来的影响；另一方面结合阅读量、讨论量、互动量与原创量等数据对样本话题影响力进行补充校验，尽量降低单一平台数据的系统性偏误。经多次筛选后，在微博平台上选取具有代表性和认知度的 35 条 AI 谣言案例作为

数据来源（见表 1）。在数据获取方面，本文采集的传播热度数据包括微博平台上样本话题阅读、讨论、转发和评论（数据截至 2024 年 7 月 25 日），以及 AI 谣言样本相关的互联网公开资料，进一步基于 35 条公开 AI 谣言提取其相关特征用以对应指标的量化判断。

表 1 AI 谣言样本示例

序号	AI 谣言	传播热度	序号	AI 谣言	传播热度
1	为和男友结婚伪造深度美颜身份证	230000000	19	四川查处两起 AI 编造传播谣言案件	921000
2	警方通报网传醉酒女子当街遭侵犯	81573000	20	女子用 AI 编造虚假地铁行凶视频被拘	657000
3	造谣公司用 AI 一天写谣言 7000 篇	44506000	21	这个 MCN 机构用 AI 编造谣言日入过万	624000
4	刘昊然被 AI 换脸造谣	21305000	22	深度伪造技术恐扰乱美国总统选举	527000
5	缅甸谣言视频为 AI 合成	15454000	23	沈泉锐粉丝 AI 造谣	474000
6	女子用 AI 制作假新闻博流量被罚	11895000	24	重庆民房爆炸消息竟是 AI 造谣	453000
7	广州限制外卖配送谣言系 AI 生成	11570000	25	男子利用 AI 造谣女婴被弃	445000
8	女子用 AI 编造校园霸凌谣言被处罚	11310000	26	网传街头斗殴竟是由 AI 生成的谣言	435600
9	男子 AI 造谣九旬老人遭遗弃成皮包骨	9832000	27	男子利用 AI 洗稿发布百万篇谣言	336000
10	AI 批量造谣水军团伙被端	9811000	28	收钱发帖一条 19800 元 2 人被批捕	335000
11	用 AI 造谣光膀男遭抱摔致死网民被处罚	9615000	29	男子用 AI 编造灾情谣言被抓捕	334000
12	利用 AI 散布谣言团伙非法获利 4 万多	8303000	30	多个卫生巾品牌致癌谣言是 AI 生成	271000
13	网民用 AI 嫁接新闻制造谣言被处罚	4034000	31	马斯克深度伪造视频骗局泛滥	218000
14	湖南一网民利用 AI 技术传播涉汛谣言	3868000	32	网传贵州女婴被弃实为 AI 生成谣言	167000
15	甘肃侦破利用 AI 炮制虚假信息案	3840000	33	自媒体用 AI 炮制 524 条假新闻被抓	166000
16	网店店主利用 AI 造谣女星离世被拘		34	美军计划用深度伪造技术传播假信息	131000
17	科大讯飞称股价下跌系 AI 虚假小作文导致	2675000	35	男子用 AI 编造民警与运钞车发生枪战谣言	49000
18	3 人用 AI 编造 3000 余个假视频被刑拘	2475000			

（二）QCA 研究方法

本文采用模糊集定性比较分析（fsQCA，fuzzy

set Qualitative Comparative Analysis）对影响 AI 谣言传播的组态关系进行研究。QCA 方法基于布尔

代数的集合论逻辑，聚焦于解决因果研究中的“组态”（Configuration）问题。即相比于单个因素各自对某结果变量的影响，QCA 更关注不同因素之间的组合关系所产生的影响作用。相比于清晰集定性比较分析（csQCA，crisp set Qualitative Comparative Analysis）对变量采取的二分法，fsQCA 允许变量在“0”到“1”之间进行取值，即评估条件程度可在“完全隶属”与“完全不隶属”之间，更适于本文相关研究变量的测量方式与取值范围。

具体操作上，QCA 步骤一般包括 5 个环节：其一，样本案例选择。QCA 对样本规模要求不高，一般适用于 15 个以上的样本案例，且通过多个案例的逻辑条件组合，在一定程度上避免了个案的特殊性对研究结论产生的干扰。其二，条件变量和结果变量确定。即根据前期研究理论和经验知识确定因果变量，并对其测量方式和赋值方法进行设计。其三，数据校准。即对测量出来的各个变量值进行校准，使其归一化到 0—1 之间。其四，通过 QCA 软件进行相关分析，包括必要性分析和组态分析，探讨影响结果变量的必要条件和组态条件。其五，对研究结果进行稳健性检验，观察子集关系、一致性和覆盖率的可能变化，以确保研究结果的可信度。

（三）变量设计与赋值

基于 QCA 构建组态模型需要基于理论或经验确定合适的条件变量和结果变量，进一步对各条件变量之间的组态关系及其对结果变量的影响作用进行探讨。一般而言，小样本规模（10—40 个样本）应将条件变量限制在 7 个左右，过多的条件变量会导致研究结果的解释复杂化。本文基于前期文献调研和行业观察，以 AI 谣言的“信源组织性”“谣言主体知名度”“内容情绪调性”“内容需求度”“媒介模态复杂度”与“受众参与度”6 个维度构建条件变量，并以“内容关注度”作为结果变量进行分析。

在确定因果变量的基础上，将各项变量拆解成若干可观测、可测量的指标因子，并基于人工编码对各指标进行赋值及加权汇总。其中 4 个一

级条件变量赋值采用四分位数赋值法，根据测量值高低依次赋值为 0、0.33、0.67、1；二级测量指标按照 0、0.5、1 进行三分位赋值；结果变量通过统计相关话题内容的阅读量测量。

此外，权重测算采用客观赋权 CRITIC 法。该方法基于评价指标的对比强度和指标之间的冲突性来综合计算指标权重，同时兼顾了指标的变异性和相关性，所计算出来的权重值越大说明该指标所包含的信息量越大、相对重要性越大。^[27] 最终确定的各项变量及其测量指标、赋值方法、指标权重如表 2 所示。

表 2 变量设计及赋值方法

变量类型	变量维度	指标分类	赋值	指标权重
条件变量	信源组织性 X1	无明确来源	0	15.75%
		个人制造	0.5	
		团伙制造	1	
	谣言主体知名度 X2	高知名度（大众知晓）	1	14.62%
		领域知名	0.5	
		不知名	0	
	内容情绪调性 X3	强冲突对立	1	17.27%
		弱冲突对立	0.5	
		强负面情感	0	
	内容需求度 X4	强需求（个人安全/利益）	1	17.68%
		弱需求（娱乐/猎奇）	0	
	媒介模态复杂度 X5	高复杂度	1	17.97%
		中等复杂度	0.5	
		低复杂度	0	
结果变量	内容关注度	受众参与度 X6	高参与度	16.70%
			低参与度	
			高关注度	100%
			中等关注度	
			低关注度	

其中, 信源组织性表示 AI 谣言发布账号的权威性、可靠性和结构化程度。谣言主体知名度指的是谣言所涉及的主体在特定群体中的知名程度。内容情绪调性指的是谣言内容中所包含的情绪色彩和情绪强度, 通过谣言内容是否包含显著的冲突和对立情绪来做判断。内容需求度指谣言内容对受众的吸引力和需求程度, 根据马斯洛的需求理论, 生理与安全为人类最基本和最强烈的需求, 故本文将强需求定义为直接关系到个人安全或利益的硬性需求,^[28] 而弱需求则为偏娱乐或猎奇的软性需求。媒介模态复杂度指的是谣言在传播过程中所使用的媒体形式和模态 (图文影音等) 的复杂程度。受众参与度指的是受众在谣言传播过程中所表现出的互动频度和二次创作程度, 具体通过评论、分享、点赞等行为来体现。此外, 内容关注度指的是用户对 AI 谣言内容的关注和传播程度。作为本研究的结果变量, 内容关注度通过 AI 谣言在微博平台上的阅读量来进行分层测量。

三、数据分析

(一) AI 谣言传播因素的结构性差异

本文对 AI 谣言的“信源组织性”“谣言主体知名度”“内容情绪调性”“内容需求度”“媒介模态复杂度”与“受众参与度”6 个维度进行数据测量, 并将各项得分以中位数 0.5 作为阈值进行分层, 认为大于等于 0.5 为高评分, 反则属于低评分。数据发现, 从各项指标综合表现来看, “内容需求度” > “媒介模态复杂度” > “内容情绪调性” > “受众参与度” > “谣言主体知名度” > “信源组织性”。

数据发现, 35 条 AI 谣言中有 68.57% 在“内容需求度”上表现出高得分 (见图 1), 可见, 当前媒介信息消费普遍功利化, 涉及个体利益的谣言信息占比更大, 也更容易激发受众非理性信任。从“媒介模态复杂度”来看, 接近六成 AI 谣言获得了较高分, 表明大多数 AI 谣言在传播时使用了较为复杂的多模态媒体形式。多模态媒体

形式 (如文字、图片、视频的结合) 能够增强信息的丰富性和吸引力。而从“信源组织性”来看, 草根传播占比较高, 意味着多数谣言的传播源头并非来自高度组织化的群体或机构, 而更多出自个人信源或无明确来源。这种低组织性可能降低了谣言的初始可信度, 但并不妨碍其在社交媒体上的传播。

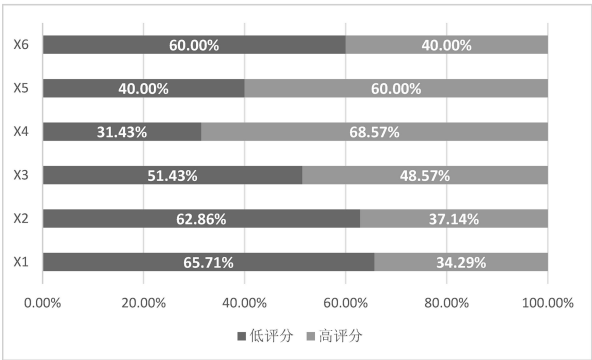


图 1 AI 谣言热度影响因素评分分布

为了进一步探讨 AI 谣言各条件变量对传播热度的影响显著性, 本文采用独立样本 *t* 检验进行分析, 结果如表 3 所示。研究发现“媒介模态复杂度”对传播热度的负相关系数为 -0.2169, *P* 值为 0.05, 达到显著水平 0.1。这表明媒介模态复杂度对谣言传播热度有显著影响, 尽管方向是负向的。这意味着高复杂度的多模态媒体形式在某些情况下降低谣言的传播热度, 可能是因为受众对复杂信息的处理负担增加, 减少了传播意愿。受众参与度对传播热度的正相关系数为 0.1988, *P* 值为 0.092, 达到显著水平 0.1。这表明受众的高参与度对谣言传播热度有显著正向影响。受众的积极互动 (如评论、分享、点赞) 能够显著提高谣言的传播热度, 增强信息的扩散效果。

其他因素如“信源组织性”“谣言主体知名度”“内容情绪调性”和“内容需求度”对谣言传播热度的影响在统计上不显著, 表明 AI 谣言的传播效果更依赖于受众的互动行为和信息呈现的复杂程度, 而非信息源的组织性或主体知名度等因素。

表 3 AI 谣言各条件变量对传播热度的影响

变量	相关系数	标准误差	<i>t</i> 值	<i>P</i> 值 ($P> t $)	显著性
X1	0.1941	0.163	1.193	0.243	不显著
X2	-0.0347	0.105	-0.33	0.744	不显著
X3	-0.0571	0.144	-0.397	0.694	不显著
X4	-0.0825	0.141	-0.585	0.563	不显著
X5	-0.2169	0.106	-2.052	0.05	显著 (显著水平 0.1)
X6	0.1988	0.114	1.744	0.092	显著 (显著水平 0.1)

此外，对 AI 谣言的传播热度分布进行分析（见图 2），谣言传播热度呈现出典型的“长尾分布”特征。少数谣言的阅读量极高，达到数千万级别，显示出强大的传播效果，而大多数谣言的阅读量相对较低，迅速下降至数百万甚至更低。这表明在 AI 谣言的传播中，少数具有高吸引力或

高关注度的谣言会迅速引起广泛关注和传播，成为热点话题，而多数谣言则只能获得有限的传播效果。长尾分布反映出信息在社交网络中传播的高度不均衡性，少数谣言能够极大地吸引受众注意力，形成爆发性的传播，而大部分谣言的传播范围较为有限。

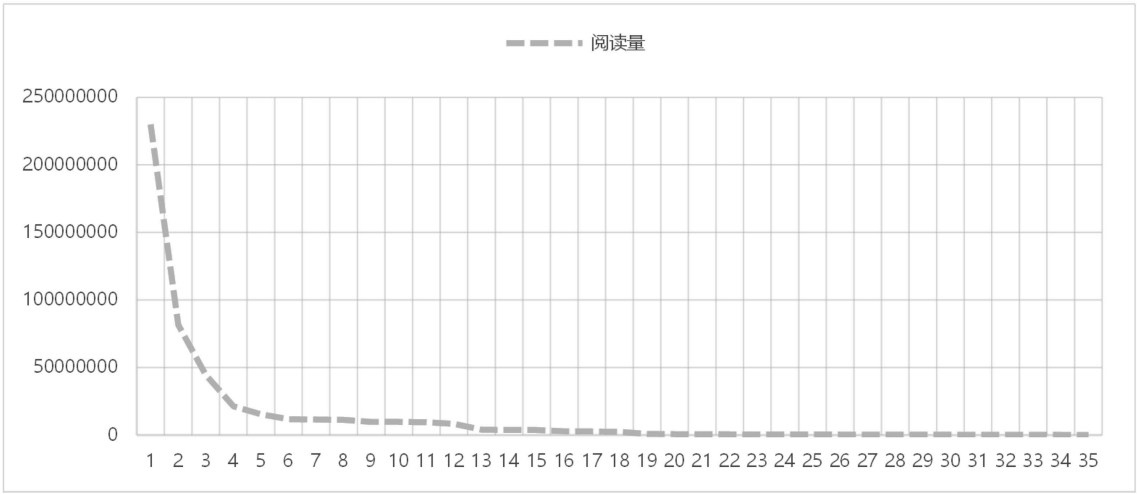


图 2 AI 谣言传播热度（阅读量）分布

由此回应了本文的研究问题一：目前 AI 谣言的“媒介模态复杂度”与“受众参与度”对传播热度的影响存在显著差异。尽管谣言的信源、内容情绪和需求度等方面的差异对传播热度的影响不显著，但复杂的多模态形式可能降低传播效果，而高参与度的谣言则能显著提升传播热度。

（二）AI 谣言大众化传播的必要性条件分析
基于 QCA 分析框架，对影响 AI 谣言传播热

度的关键因素进行单变量分析。其中，以“信源组织性”“谣言主体知名度”“内容情绪调性”“内容需求度”“媒介模态复杂度”与“受众参与度”作为条件变量，“内容关注度”作为结果变量。通过模糊集定性比较分析 fsQCA 软件分析各条件变量对结果变量的解释程度和充分必要性，当变量的一致性指标大于 0.8 时，该条件变量可视为结果变量的充分条件，而当一致性大于 0.9

时，可认为该条件变量为结果变量的必要条件。

通过表 4 分析结果可以发现，对于谣言传播热度这一结果变量，“信源组织性”高评分对高热度的必要性一致性较高（0.8414），覆盖度为 0.5031，表明信源组织性高对高传播热度有较强

的解释力，但覆盖面有限。“内容需求度”高评分对高热度的一致性较高（0.8297），表明高需求度对高传播热度有较强的解释力。“信源组织性”与“内容需求度”的一致性大于 0.8，构成 AI 谣言传播热度的充分条件。

表 4 必要条件分析结果

结果变量/条件变量			高热度		低热度	
			一致性	覆盖度	一致性	覆盖度
AI 谣言传播热度	X1	高评分	0.8414	0.5031	0.6518	0.6809
		低评分	0.4662	0.4339	0.5243	0.8524
	X2	高评分	0.3940	0.3719	0.3989	0.6578
		低评分	0.6374	0.3777	0.6190	0.6409
	X3	高评分	0.7527	0.4191	0.6968	0.6779
		低评分	0.4215	0.4431	0.4030	0.7401
	X4	高评分	0.8297	0.3924	0.7911	0.6537
		低评分	0.2677	0.4231	0.2646	0.7308
	X5	高评分	0.4419	0.2681	0.6905	0.7319
		低评分	0.5581	0.5079	0.3095	0.4921
	X6	高评分	0.5612	0.5107	0.3077	0.4893
		低评分	0.4388	0.2662	0.6923	0.7338

由此回应了本文的研究问题二：目前 AI 谣言在传播热度层面，尚不存在必要性因素。即 AI 谣言的传播热度更可能是多种因素共同作用的结果，而不是单一属性所能决定的。政策制定者和平台管理者在应对 AI 谣言时，应采取多层次、多角度的综合措施，以有效遏制谣言的扩散。

（三）AI 谣言大众化传播的组态条件分析

进一步对 AI 谣言各条件变量构建真值表并进行组态分析，以探讨构成 AI 谣言传播热度的最优条件组合。本文针对“信源组织性”“谣言主体知名度”“内容情绪调性”“内容需求度”“媒介模态复杂度”与“受众参与度”6 个一级社会线索作为条件变量进行组态分析，结果如表 5 所示。

表 5 高热度 AI 谣言的最优组态分析

条件组态	组态 1	组态 2	组态 3	组态 4
X1	○	●	○	○
X2	○	●	○	●
X3	●	●	○	●
X4	●	○	○	●
X5	○	○	○	●
X6	●	○	●	●
简单解一致性	0.8267	1.0000	1.0000	0.6313
简单解覆盖度	0.2920	0.1185	0.1601	0.0981
中间解一致性	1.0000	1.0000	1.0000	1.0000
中间解覆盖度	0.1201	0.0400	0.0400	0.0400

数据发现,覆盖率最高的中间解为:低信源组织性*低主体知名度*高情绪调性*高内容需求度*低媒介模态复杂度*高受众参与度(覆盖度 0.1201)。这一中间解表明,当 AI 谣言同时具有极端的情绪(如愤怒、恐慌)和大众利益关联性——即内容涉及个人安全或利益,并且充分调动受众二次传播和互动时,其传播热度最为显著。这也反映了 AI 谣言的“弱势信息增强模式”,其初始条件并不具备明显的优势,但通过充分调动大众的情绪和参与度,通过社交网络的自组织机制,使得谣言信息传播得以增强和放大。这种模式显示了在网络环境中,即使是源自不知名主体、结构简单的信息,也可以通过受众的集体互动和“长尾效应”实现非线性增长。

此外,高信源组织性*高主体知名度*高情绪调性*低内容需求度*低媒介模态复杂度*低受众参与度(覆盖度 0.04)也构成了 AI 谣言传播的一种组态路径。这类谣言一般具备组织性、规模性、针对性,虽然单条信息的受众参与度不高,与大众直接利益关联度不强,但往往针对某一特定对象或知名主体进行“蓄谋”性造谣,往往存在资本操纵嫌疑。

由此回应了研究问题三:高传播热度的 AI 谣言普遍需要具备强烈的内容情绪调性、高内容需求度和高受众参与度。草根发布的“强情感”“高需求”信息和组织性发布的“强冲突”“强针对性”信息为两种典型模式。

四、研究结论与讨论

通过对 35 个 AI 谣言传播案例的定性比较分析,本研究在理论层面揭示了谣言发布主体、主题、内容、模态、受众等影响因素对其传播热度的组态影响。基于 5W 传播框架——谁发布、发布什么内容、什么形式发布、面向受众是谁、传播效果如何,系统性梳理 AI 谣言传播的内在机制。在实践层面,本研究的发现为 AI 谣言的预防和控制提供了实证支持,通过对谣言属性进行多维识别和组态计算,有助于实现前瞻性风险研判和精细化辟谣管理。诚然, AI 谣言的传播机制复杂且多样,各属性之间的组态效应与交互作用 and 对其跨域、跨圈、跨平台传播的影响需在未来研

究中进一步细化探索。

(一) AI 谣言传播热度的三级断层:“二八效应”凸显

本文发现热度位于前 20% 的 AI 谣言占据了超过八成(84.70%)的舆论声量。头部谣言通常具有高度情绪化的内容表达,并借助于 AI 生成技术的视觉增强效果,放大了其传播的渗透力和多平台触达能力。相比之下,中部谣言和长尾谣言则表现出较为显著的传播断层,前者在一定程度上通过针对特定受众的定向传播获得了较为稳定的关注度,而后者则更多依赖于特定垂直社群的纵向传播动力。

这种现象也映射了 AI 技术在谣言制造中的“量产化”趋势——技术门槛的降低和生成工具的普及使得大量相似度高的谣言能够在短时间内批量生成并广泛传播。尤其是在社交媒体平台上,算法推荐机制加剧了这些头部谣言的传播扩散,进一步强化了“二八效应”的现实表现。与此同时,信息审查和辟谣机制的升级也使得更多谣言在传播早期被及时遏制,虽然在大众传播场域热度较低。有不少谣言通过特定的社群或私域流量进行圈层化传播,尽管其影响力相对有限,仍在局部范围内产生了负面的社会效应。这些谣言通常针对特定人群,以较为隐蔽的方式传播,难以被快速识别和遏制。它们在特定群体中维持着一定程度的关注度,并通过不断重复和渲染,逐步深化了错误信息对公众认知的侵蚀。此外,大量“长尾谣言”的存在也进一步复杂了谣言治理的挑战。虽然其传播热度较低,但其程式化的生产方式和规模性的发布模式却形成了持续性的信息污染。

(二) AI 谣言传播四大要素:内容 > 媒介 > 受众 > 信源

相比于传统信息传播中对信源权威性、可信度的强依赖, AI 谣言的传播更多依赖于内容的感染力和媒介视觉呈现效果。技术赋能下,大量非专业的草根信源也能制造出高度真实感的内容,“眼见为实”带来的认知偏差弱化了受众对信源的感知能力,进而导致了信息真伪的判断更加依赖于直观印象而非理性分析。在这种环境下,谣

言的传播机制表现出明显的视觉导向性, 强调内容的情绪渲染和媒介的感官刺激。

根据认知心理学, 人类认知的深度从低级到高级可以划分为感觉、知觉、表象、概念、判断、推理六大层次, 进一步可以抽象为包括神经、心理层面的“低阶认知”和包括语言、思维和文化层级的“高阶认知”^{[29][30]}。从这个层面来看, AI 谣言对受众的影响主要集中在低阶认知层面, 而辟谣的关键则在于高阶认知的激活。这需要公众不仅接受信息的直接呈现, 而且能够对信息进行更深层的批判性思考和逻辑分析。在视觉和情绪的强烈刺激下, 触发低阶认知的谣言很容易获得快速传播, 因为这类认知过程更直接、更本能。如何通过人机协同的模式, 让受众在构建感性认知模式的同时, 通过自动化的风险要素识别和提示激活用户的理性评价机制, 是未来 AI 谣言治理中需重点探索的策略。

(三) AI 谣言传播两种典型模式: 草根创作的“强情感”> 权威背书的“强冲突”

通过 QCA 组态分析发现, “强情感”和“强冲突”是目前 AI 谣言大众化传播的两种典型模式, 前者更多以流量为导向, 通过刺激大众化情绪、激活二次传播来实现注意力收割, 参与主体多为草根自媒体; 而后者更多存在资本操纵和政治内嵌, 针对性更强, 通常存在组织性操作, 部分会通过所谓的“权威联动”来增强谣言的可信度。“强情感”模式的谣言通常包含戏剧化元素, 利用恐惧、愤怒或同情等基本情绪, 夸大未经验证的“事实”或编造符合受众“固有认知倾向”的内容来激发受众的共鸣, 促进信息的扩散。而“强冲突”模式则依赖于权威背书或似是而非的逻辑来构建信息的冲突性, 这一模式在政治、企业竞争或社会议题相关的谣言中尤为常见, 其中信息往往涉及政策争议、社会不公等引发广泛讨论的话题。

(四) AI 谣言治理的破局痛点: 多模情绪交织与大众二次创作

从 AI 谣言的传播热度影响因素来看, “媒介模式复杂度”和“受众参与度”对谣言传播形成显著影响。尤其是对于强情绪渲染的谣言内容而

言, 视觉框架和情绪符号的使用大幅提升了谣言的感染效率, 提高了受众信息接受和二次传播动机。同时, 各类开源技术的应用使大众参与信息的创作和再创作变得更为容易, 这一过程中, 原始信息往往经过多次修改、扭曲, 甚至完全脱离最初的事实基础, 形成全新的内容实体。这种由下而上的内容生成机制, 极大地复杂化了信息真伪的辨别过程, 也使得谣言治理工作面临前所未有的挑战。在 AI 增强的信息流转中, 多层级的信息交互被显著加速, 同时谣言的修正与抵制难度也相应增大。多模态信息的综合运用在刺激受众情绪反应与促进信息共享方面显示出其固有的优势。这种技术环境下的受众转变为信息的主动参与者和再创作者, 使得谣言在不同群体中衍生出不同“变体”, 甚至形成“千人千面”的针对性传播, 每一次信息的重构和分享都有可能使谣言进一步变异, 更深地植入受众群体之中, 也使得 AI 谣言的可持续治理更加任重道远。

本文仍存在部分问题有待进一步优化与探讨。首先, 在采样上考虑到不同平台的传播特性差异和数据量级差异, 选择了微博作为 AI 谣言传播的取样平台, 在样本规模上存在一定局限性, 相关分析结论是否适用于其他平台有待进一步验证。其次, 限于数据可获取性和标注难度, 本文将影响谣言的传播变量简化为 6 大维度, 并据此设计量化因子, 在一定程度上忽略了 AI 谣言在视觉层面的可信度、在传播层面的恶意推广和资本投入等要素。再次, 本文选择谣言的阅读量作为结果变量, 弱化了 AI 谣言实际产生的社会影响, 相关指标有待进一步丰富。最后, 由于 AI 谣言传播具备较强的隐蔽性, 本文所研究的对象主要是“已被辟谣”的 AI 谣言, 仍有部分未能被明确识别的研究样本还需挖掘。如何从更深层次去挖掘 AI 谣言的隐性特征, 也成为后续研究需要进一步关注和探讨的问题。

参考文献:

[1] Paschen, J. (2020). Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product & Brand Management*, 29 (2): 223

- 233.
- [2] Vaccari, C. & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media+ Society*, 6 (1): 1-13.
- [3] 郭庆光. 传播学教程 [M]. 北京: 中国人民大学出版社, 1999: 96-98.
- [4] 刘建明. 舆论传播 [M]. 北京: 清华大学出版社, 2001: 291.
- [5] 江万秀, 雷才明, 江凤贤. 谣言透视 [M]. 北京: 群众出版社, 1991: 15.
- [6] [法] 让诺埃尔·卡普费雷. 谣言: 世界最古老的传媒 [M]. 郑若麟, 译. 上海: 上海人民出版社, 1991: 15-17.
- [7] 胡泳. 谣言作为一种社会抗议 [J]. 传播与社会学刊, 2009 (9): 67-94.
- [8] 李永平. 文学人类学视野下的谣言、流言及叙述大传统 [J]. 思想战线, 2014 (2): 23-29.
- [9] Hajli, N., Saeed, U., Tajvidi, M. & Shirazi, F. (2022). Social bots and the spread of disinformation in social media: the challenges of artificial intelligence. *British Journal of Management*, 33 (3): 1238-1253.
- [10] Xu, J., Xian, L., Liu, Z., Chen, M., Yin, Q. & Song, F. The future of combating rumors? Retrieval, discrimination, and generation. Retrieved May 28, 2024, from <https://arxiv.org/pdf/2403.20204>.
- [11] 李玉碧, 胡曙光. 人工智能时代网络谣言的嬗变: 基于自媒体造谣朱婷的案例分析 [J]. 中国传媒科技, 2022 (12): 25-28.
- [12] Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9 (11): 39-49.
- [13] [美] 奥尔波特, 等. 谣言心理学 [M]. 刘水平, 等, 译. 沈阳: 辽宁教育出版社, 2003: 3.
- [14] 喻国明. 网络谣言的文本结构与表达特征——基于腾讯大数据筛选鉴定的 6000+ 谣言文本的分析 [J]. 新闻与写作, 2018 (2): 53-59.
- [15] 罗自文, 林燕飞. 眼见为虚: 健康类短视频谣言的叙事特征与社会心理——基于今日头条谣言数据库的实证研究 [J]. 中国社会科学院研究生院学报, 2021 (6): 93-103.
- [16] 彭柳, 陈红飞. 突发性重大公共卫生事件网络谣言文本特征及治理——以中国互联网联合辟谣平台所辟新冠疫情谣言为例 [J]. 华南师范大学学报 (社会科学版), 2021 (2): 183-192, 208.
- [17] 朱梦蝶, 付少雄, 郑德俊, 等. 文献视角下的社交媒体健康谣言研究: 特征、传播与治理 [J]. 图书情报知识, 2022 (5): 131-143.
- [18] Hameleers, M., Powell, T. E., Van Der Meer, T. G. & Bos, L. (2020). A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. *Political Communication*, 37 (2): 281-301.
- [19] 丁学君, 刘思奇. 基于演化博弈的在线社会网络谣言传播行为研究 [J]. 运筹与管理, 2020 (11): 11-21.
- [20] 宁珊, 熊晶, 贺文晨. 人工智能背景下的网络谣言治理研究 [J]. 中国信息安全, 2023 (8): 73-76.
- [21] 刘超民. 生成式人工智能场景下虚假信息风险特殊性透视及应对 [J]. 中国海洋大学学报 (社会科学版), 2024 (2): 112-121.
- [22] 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 (3): 112-120.
- [23] Liu, S. (2020). Countermeasures of internet rumor management based on artificial intelligence technology. In Stephan, R. (eds.). *Journal of Physics: Conference Series*. Bristol: IOP Publishing, 1-6.
- [24] 黄河. 网络谣言的智能化演变及治理 [J]. 人民论坛, 2023 (4): 62-65.
- [25] Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S. & Sun, L. A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. Retrieved May 28, 2024, from <https://arxiv.org/pdf/2303.04226>.
- [26] Oakley, D. P. & Rogg, J. (2024). Spreading the “smog of war”: the impact of propaganda, social media, and OSINT on US civil-intelligence relations. *Intelligence and National Security*, 39 (3): 539-553.
- [27] 王昆, 宋海洲. 三种客观权重赋权法的比较分析 [J]. 技术经济与管理研究, 2003 (6): 48-49.
- [28] 胡家祥. 马斯洛需要层次论的多维解读 [J]. 哲学研究, 2015 (8): 104-108.
- [29] 蔡曙山. 认知科学框架下心理学、逻辑学的交叉融合与发展 [J]. 中国社会科学, 2009 (2): 26-39.
- [30] 蔡曙山. 论人类认知的五个层级 [J]. 学术界, 2015 (12): 9-24.