

理性的缺失和回归——一项关于大学生微信群聊中社交机器人传播谣言的实验研究

贾钊昊

澳门科技大学人文艺术学院 中国澳门

【摘要】本研究旨在探究浏览社交机器人生成的谣言及其评论对大学生微信用户愤怒情绪、恐惧情绪、转发行为的影响，以及平台法律风险提示的作用。在广州大学城于 2024 年 10 月开展控制实验，通过多种渠道从不同高校招募实验对象(N=160, 年龄 18-26 岁, 女生 50%)。结果表明，浏览包含情绪化评论内容的谣言会显著提升用户转发的概率，愤怒情绪和恐惧情绪在此过程中具有中介作用。此外，平台法律风险提示可以显著降低用户转发谣言的概率，尽管该因素对谣言评论、负面情绪、转发行为这三者关系的调节作用并不显著。这项研究深化了对社交机器人在大学生微信群聊中传播谣言机制的理解，并从政策制定者、社交平台、用户个人三个层面为应对人工智能所生成与传播的网络谣言提供了建议。

【关键词】社交机器人；网络谣言；大学生微信群聊；控制实验

【收稿日期】2025 年 5 月 7 日

【出刊日期】2025 年 6 月 5 日

【DOI】10.12208/j.ssr.20250212

The absence and return of rationality: An experimental study on social bots spreading rumors in university WeChat groups

Zhaohao Jia

Faculty of Humanities and Arts, Macau University of Science and Technology, Macau

【Abstract】 This study aims to explore the impact of viewing rumors and their comments generated by social bots on college students' anger, fear, and sharing behaviors in WeChat, as well as the role of platform issued legal risk warnings. A controlled experiment was conducted in Guangzhou University Town in October 2024, recruiting participants (N=160, age 18-26, 50% female) from various universities through multiple channels. The results show that viewing rumors with emotional comments significantly increases the likelihood of users sharing them, with anger and fear serving as mediating factors in this process. Additionally, platform issued legal risk warnings can significantly reduce the likelihood of users sharing rumors, although this factor does not significantly moderate the relationship between rumor comments, negative emotions, and sharing behaviors. This study advances our understanding of how social bots spread rumors in university WeChat groups and offers suggestions for policymakers, social platforms, and individual users to address online rumors generated and disseminated by artificial intelligence.

【Keywords】 Social bots; Online rumors; University WeChat groups; Controlled experiment

随着人工智能技术的发展，社交机器人（Social bots）在互联网信息传播中扮演着愈发重要的角色，他们逐渐变得智能且高效，能够通过模仿人类用户的行为在社交网络平台上迅速扩散信息。与具备物理形态、设计用于和人类进行互动的社交机器人（Social robots）不同，本文探讨的社交机器人是指一类算法驱动的程序，用于模拟人类行为在社交媒体上自动发布内容、参

与用户互动和传播信息。研究表明在有意操纵下，社交机器人可通过混淆真假信息、使用标签策略和渲染虚假态势等手段，在认知传播各环节中发挥关键作用，从而推动阴谋论叙事、煽动反智情绪、加剧意见极化和虚假信息的传播^[1]。这种现象在新冠疫情期间尤为明显，社交机器人被发现大量参与涉疫叙事的生产和传播，影响公众对疫情态势的感知。此外，伪科学信息的扩散

作者简介：贾钊昊（1999-）男，澳门科技大学人文艺术学院传播学博士研究生，广东外语外贸大学新闻与传播专业硕士，研究方向为媒介心理学、社交媒体量化研究。

对网络科学传播和科普工作开展形成了挑战, 优质科普内容的传播渠道受阻且难以引起公众关注, 而商业营销组织则利用伪科学概念从中获利, 尤其是在健康民生领域^[2]。

在国际层面, 伪科学和虚假信息的传播已成为大国间博弈的一种手段, 对意识形态安全和社会稳定构成了潜在威胁^[3]。

研究社交机器人在大学生微信群聊中传播谣言的影响因素及其机制具有重要意义。在理论层面上, 研究机器人如何模仿人类交流方式来增强谣言的可信度, 可以加深对个体在复杂信息环境中决策机制的理解, 这有助于解释为什么某些谣言在特定条件下更容易扩散。此外, 随着 AIGC 技术的发展, 关于社交机器人与人类互动的研究为“技术——人类交互”理论提供了新视角, 研究社交机器人如何在微信群中与大学生互动并推动谣言传播, 能够拓展人们对技术中介及其互动模式的理解, 深化对于人工智能时代技术风险的认识。

在现实层面上, 大学生作为一个高度活跃、互动频繁的群体, 社交平台谣言对他们的影响通常更为严重, 且更容易在群体内发生二次扩散。研究社交机器人如何在大学生微信群中生成和传播谣言, 有助于开发有效的应对策略, 特别是在虚拟环境中如何设计更高效的防御性传播机制。最后, 这项研究还能为大学生的媒介素养教育提供经验支持, 通过深入分析社交机器人传播谣言的影响因素及作用机制, 可为大学生群体设计出更具针对性的数字教育策略, 提升他们对社交平台信息的甄别与批判能力。

1 文献综述与假设提出

1.1 “知-信-行”行为理论

“知-信-行”行为模式 (KAP) 是一种用于理解行为产生过程的理论模型。它将行为产生分为三个连续阶段, 即获取知识 (Knowledge)、产生信念 (Attitude)、形成行为 (Practice)。获取知识阶段是行为产生的基础, 通过学习和掌握相关信息, 人能够对某一事物的利弊形成认识。在产生信念阶段, 对事物的主观认识会转化为个体的信念和态度, 这也是行为产生的直接动力, 当人们对所掌握信息进行思考, 便会唤起强烈的情感进而转化为信念。形成行为阶段是指信念和态度最终促使个体采取实际行动, 只有当主观认识上升为信念, 人才会采取相应措施并调整行为策略。该模型广泛应用于心理与教育领域, 通过调查目标人群的知识、信念和行为情况, 进行特定的干预效果评估^[4]。

在个体层面上, KAP 理论可帮助理解网络谣言的

扩散规律, 包括信息获取、情感形成和传播行为三个阶段。网络谣言的传播往往利用人对某些信息的缺乏或误解, 其内容通常涉及公众关注的热点话题, 但相关的真实信息可能不易获取或带有片面性, 这种信息不对称性为谣言的传播提供了土壤^[5]。当互联网用户接触到谣言时, 他们的态度和信念很容易受到影响, 因为谣言往往通过煽动情感和强化偏见来影响人们的信念, 使他们更容易相信并传播这些未经证实的信息, 尤其是涉及健康、经济或安全的内容更容易唤起用户的愤怒和恐惧情绪, 从而增强其传播力^[6]。而信念的改变最终会影响人的行为, 相信谣言的用户可能会采取特定行动, 如分享转发、参与讨论等, 这些行为将进一步扩大谣言的传播范围和影响力^[7]。

综上所述, 本文以 KAP 理论为框架搭建研究模型, 在此基础上针对网络谣言的相关评论、个体愤怒情绪、个体恐惧情绪、转发风险提示等变量之间关系提出假设, 探讨社交机器人在大学生微信群聊中传播谣言的影响因素和规律。

1.2 网络谣言对用户情绪的影响

在大多数情况下, 网络谣言会对用户情绪造成负面影响。一方面, 网络谣言中所包含的夸张和虚假信息很容易引发用户的恐惧和焦虑情绪, 在重大突发事件或公共安全事件中, 谣言会导致用户对自身安全产生担忧, 加重人与人之间的不信任感, 而个体的认知与情绪变量、群体变量、情境变量都会对谣言传播行为造成影响^[8]。另一方面, 当用户发现自己既有的态度倾向与认知受到挑战时, 常常会产生愤怒和厌恶情绪, 这种情绪反应对于涉及道德或法律问题的谣言尤为明显, 是影响其后续传播的重要因素^[9]。当持续接触到负面的谣言内容后, 用户还可能会感到紧张和压力, 这会对其日常生活和心理健康造成不良影响。而愤怒与恐惧等负面情绪在影响用户心理健康的同时还可能致谣言的进一步扩散, 因为情绪化的谣言能通过情绪感染使受众产生类似的情绪反应, 从而增加其对谣言的传播意愿^[10]。

基于上述, 本文选取愤怒与恐惧这两种在谣言传播中常见的用户情绪作为观测变量, 考察在不同条件下社交机器人所生成的谣言及其相关内容对大学生用户情绪以及后续转发行为的影响。

1.3 社交媒体评论对用户情绪的影响

社交媒体上来自他人的评论内容会对用户情绪产生显著影响。用户在阅读评论时很容易受到他人情绪的影响, 尤其是负面情绪如愤怒、恐惧、悲伤等, 这种

情绪感染机制会导致用户自身情绪发生波动和变化^[11]。而且受到情绪放大效应的影响,社交媒体平台上的负面情绪更容易广泛传播,也更容易引发用户的非理性行为,包括群体模仿与盲目跟从等。此外,社交媒体上的评论常常会引发情绪极化现象,即用户的情绪和观点在阅读相关评论后变得更加极端,此现象在涉及敏感话题时尤为明显,而负面的评论内容还会增加用户的心理压力,引发恐惧和焦虑等心理问题^[12]。因为在后真相时代,理性思考往往让位于情绪宣泄,这种现象导致网络舆论场呈现出两极对立的状态,即网民的意见观点集中在两个极端,中间理性客观的观点表达则相对较少,在沉默的螺旋机制下非理性的意见表达最终会占据主流。

在人工智能时代,结合深度伪造、聊天机器人等技术,谣言及其评论内容变得更加难以辨别且容易引起公众更强烈的情绪反应。综上所述,本文提出以下假设:

H1.在浏览社交机器人带有情绪的评论后,谣言内容会唤起用户更高的愤怒情绪。

H2.在浏览社交机器人带有情绪的评论后,谣言内容会唤起用户更高的恐惧情绪。

1.4 谣言引起的负面情绪对用户转发行为的影响

网络谣言引起的愤怒、恐惧等负面情绪可显著影响用户的转发意愿。研究表明,情绪化的谣言能够引发受众产生类似的情绪反应,在情绪感染机制作用下用户很容易将自身代入事发情境从而产生共情,并增加谣言的转发概率^[13]。

在负面情绪的影响下,用户往往缺乏对信息理性分析的能力,更容易相信并转发未经证实的信息,这种认知偏差使谣言得以在网络上迅速扩散。此外,当用户感受到强烈的负面情绪时,他们会倾向于通过分享与转发来表达自己的情感和立场,从而获得群体内部的认同和支持,这与个体的基本心理需求相关^[14]。再者,负面情绪还可能引发“逆火效应”,即当用户面对与其既有信念相冲突的信息时,其原有观点不仅不会改变,反而会变得更为坚定,从而进一步推动谣言的传播^[12]。最后,感知风险也能够调节负面情绪对转发意愿的影响,对身边环境过高的风险感知会增加用户的不安情绪以及恐惧心理,通常在缺乏安全感状态下用户会更容易相信并传播谣言^[15]。

这些因素的共同影响,使负面情绪在网络谣言传播中起到重要的助推作用。综上所述,本文提出以下假

设:

H3.用户的愤怒情绪水平正向影响其转发谣言的数量。

H4.用户的恐惧情绪水平正向影响其转发谣言的数量。

H5.在社交机器人评论内容对用户谣言转发行为的影响路径上,愤怒情绪具有中介作用。

H6.在社交机器人评论内容对用户谣言转发行为的影响路径上,恐惧情绪具有中介作用。

1.5 谣言后果联想对用户转发行为的影响

用户对行为后果的联想会影响其转发与分享网络谣言的意愿。当用户意识到传播谣言可能带来负面后果时,包括不良社会影响、道德法律风险等,他们的传播意愿会明显降低。例如在健康谣言的传播中,添加警告信息可以有效减少用户的转发意愿,其中来自于行业专家警告信息的干预效果最为理想,这反映出互联网传播模式下意见领袖在两极传播中的作用机制^[16]。此外,个体在接触谣言信息后的情绪状态和所选择的认知加工模式也会影响其行为决策,在核心路径分析模式下用户会更多地进行理性思考,对各种后果的可能性进行模拟、分析、评估,从而帮助自身在实现期望效益最大化的同时尽量规避风险,反之在边缘路径处理模式下用户更有可能因为冲动而做出非理性的行为决策^[17]。

可见,现有研究已证实来自其他用户的警告信息可以减少用户转发谣言的意愿,然而来自社交平台的法律责任风险提示对用户谣言传播行为的干预效果尚未得到有效评估。基于上述,本文提出以下假设:

H7.来自社交平台的法律风险提示可降低用户转发谣言的频率。

H8.在愤怒情绪对用户谣言转发的影响路径上,法律风险提示具有负向调节作用。

H9.在恐惧情绪对用户谣言转发的影响路径上,法律风险提示具有负向调节作用。

2 研究设计

2.1 研究框架

通过对相关文献与研究成果的梳理,本文共罗列出5个变量,包括社交机器人对谣言的评论、个体的愤怒情绪、个体的恐惧情绪、平台的法律风险提示、用户的扩散与转发,并依照前文所提出研究假设的逻辑结构与顺序关系构建出总体研究框架(如图1)。

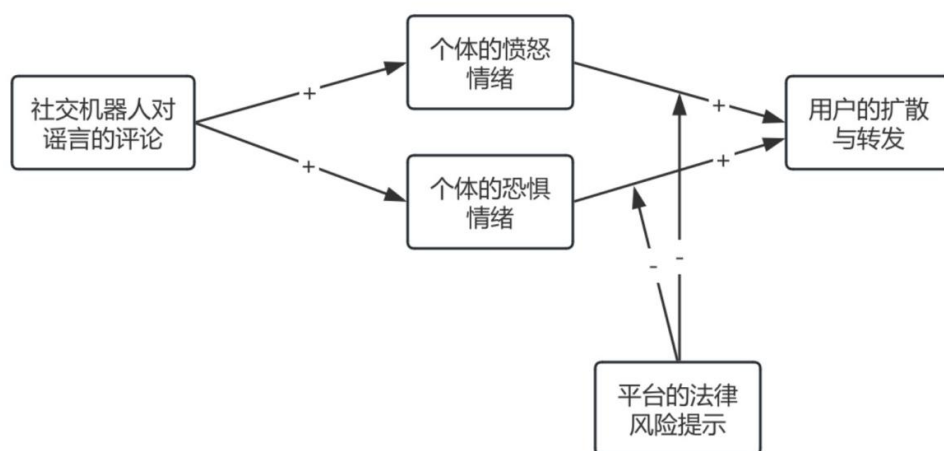


图1 研究模型 (Research Model)

2.2 抽样程序与实验准备

本研究选址地点为广州番禺大学城，通过校友互助群、微信校园墙招募实验对象，并在完成整个实验流程后向他们支付一定金额的酬劳。按照实验设计，所招募志愿者将被分为4组，包括含评论含提示组、含评论不含提示组、不含评论含提示组、不含评论不含提示组。为保障样本的广泛性和代表性，每组都需要40名志愿者，他们分别来自于大学城10所高校，即每所高校招募4人，其中包括男本科生1名、男研究生1名、女本科生1名、女研究生1名。所有组别均按照这个定额进行配置，共招募160人参与本次实验。被招募的志愿者在实验开始前会浏览研究的事先声明，并签署知情同意书以及隐私保护协议。

在招募志愿者时设置了两道筛选程序。第一，被招募者必须是微信用户且日常有查看群聊消息的习惯。第二，被招募者还需通过简要的心理状况与认知功能评估，设置该程序的目的是为了减少前期无关因素对实验结果的影响，只有心理状况与认知功能均正常的志愿者才能正式成为实验对象。

通过筛选程序后，参与者成为实验对象并被随机分配到不同组别中。以来自中山大学的4名男本科生为例，每位实验对象会抽签获得一个按自然数顺序排列的编号，即“1”、“2”、“3”、“4”，对男研究生、女本科生、女研究生执行相同操作，其他高校也以此类推。为保证实验分组随机化的成功，将编号为“1”的志愿者分到含评论含提示组，将编号为“2”的志愿者分到含评论不含提示组，将编号为“3”的志愿者分

到不含评论含提示组，将编号为“4”的志愿者分到不含评论不含提示组。

根据大学校园生活场景，运用生成式人工智能制作了4条谣言内容，涵盖了个人绯闻、食品安全、传染疾病、校园犯罪这四种常见的谣言类型，再针对每条谣言生成若干不含额外事件信息的主观情绪评论，最后通过社交机器人模拟出微信群聊场景。此外，利用生成式人工智能编写了一段关于用户转发行为的法律责任提示，并将其内嵌到模拟的群聊场景中，当用户按下“转发”选项时自动触发，通过微信群聊风险提示的形式呈现。

2.3 预实验

为保障实验刺激的有效性，在正式实验开始前开展预实验，共招募10名大学生志愿者（这些人不会出现在正式实验中）。提前准备好4则谣言内容及相关用户评论，每条内容安排两名志愿者负责，其中一人仅阅读谣言内容，另一人在浏览谣言内容后还查看了相关评论。操作结束后与参与者进行面对面访谈，了解每条内容及其评论是否能引起他们的愤怒和恐惧。对于情绪唤起效果不及预期的材料，采用生成式人工智能进行修改，并重新招募两名志愿者重复上述过程，直到刺激效果达到实验标准。最后剩余两名志愿者，其中一人会先被告知内容转发的法律责任风险，然后进行模拟微信群聊，最后通过访谈来了解其转发行为意愿，而另一人则在不被告知法律责任风险的情况下进行模拟群聊和访谈。了解他们的转发行为意愿是否与法律责任风险的知悉相关，并根据结果对风险提示作出调整

和修改。

事后访谈结果显示,相较于仅浏览谣言内容,查阅谣言及其附带评论的被试对象均表现出更强烈的情绪波动。此外,被告知内容转发法律风险的被试在群聊互动与内容分享的环节上表现得更为冷静且克制。以上表明实验刺激的设计有效,不需要做出进一步改动。

2.4 正式实验

为尽可能排除干扰因素的影响,正式实验选在安静的空教室中进行,通过与实验对象提前沟通确定好具体时间和地点。实验开始前,所有被试都需要将手机设为静音后放到讲台上,然后进行 5 分钟的情绪调整与精神放松,保证接受实验刺激前所有人都处于平静状态。

进入实验后,含评论含提示组的被试先浏览第一则谣言内容及其对应的评论,浏览完毕后需要回答一道与内容相关的问题作为操纵检查,然后对自身的愤怒与恐惧情绪作出自我评估,采用 5 点李克特量表(“上述内容使我感到愤怒”、“上述内容使我感到恐惧”; 1=完全不同意, 5=完全同意)。随后进入到内容转发界面,让被试假设转发对象是其任一好友,可选择“转发”或“取消”选项。若点击“转发”选项,则会出现法律责任风险提示的界面,并在下方提供“继续”和“取消”选项供被试选择,对选择“继续”的进行转发计数。对剩余三则谣言内容重复执行以上操作,最后统计出每位被试的平均愤怒水平、平均恐惧水平、转发总数。

对于含评论不含提示组,采用相同的内容排列顺序。被试先浏览第一则谣言内容及其对应的评论,浏览完毕后接受相同的操纵检查程序,然后对自身的愤怒与恐惧情绪作出评估。随后进入转发界面,同样让被试假设转发对象是其好友,可选择“转发”或“取消”选项,若点击“转发”选项,则直接作为转发计数。对剩余三则谣言内容重复执行以上操作,最后统计出每位被试的平均愤怒水平、平均恐惧水平、转发总数。

对于不含评论含提示组,采用相同的内容排列顺序。被试先浏览第一则不包含评论的谣言内容,浏览完毕后接受操纵检查,然后对自身的愤怒与恐惧情绪作出评估。随后进入转发界面,同样让被试假设转发对象是其好友,可选择“转发”或“取消”选项,若点击“转发”选项,则会出现法律责任风险提示的界面,并在下方提供“继续”和“取消”选项。对剩余三则不包含评论的谣言内容重复执行以上操作,最后统计出每位被试的平均愤怒水平、平均恐惧水平、转发总数。

对于不含评论不含提示组,采用相同的内容排列顺序。被试先浏览第一则不包含评论的谣言内容,浏览完毕后接受操纵检查,然后对自身的愤怒与恐惧情绪作出评估。随后进入转发界面,同样让被试假设转发对象是其好友,可选择“转发”或“取消”选项,对选择“转发”的被试进行转发计数。对剩余三则不包含评论的谣言内容重复执行以上操作,最后统计出每位被试的平均愤怒水平、平均恐惧水平、转发总数。

实验结束后,所有参与者需要向研究人员提供性别、年龄、学习阶段的信息,并告知其先前浏览的所有内容以及聊天场景均为 AI 虚构,以防对他们的现实生活造成不良影响。

3 数据分析与假设检验

正式实验共有 160 名被试参与,其中每组 40 人,总共 4 组。采用 SPSS 29.0 软件和 Process 插件程序对实验数据进行分析,从而检验各假设成立情况。

3.1 数据信度

对所有被试愤怒情绪与恐惧情绪的自我评估数据进行可靠性分析(见表 1)。结果显示,对 4 则谣言的愤怒自评总信度系数 Cronbach alpha 为 0.855,且对各项进行逐一剔除后的 Cronbach alpha 均小于 0.855;对 4 则谣言的恐惧自评总信度系数 Cronbach alpha 为 0.828,且对各项进行逐一剔除后的 Cronbach alpha 均小于 0.828,说明实验数据的可靠性良好。

表 1 可靠性分析 (Reliability Analysis)

变量	题项	删除项后的 Cronbach alpha	Cronbach alpha
愤怒情绪	对第一条谣言的愤怒	0.812	0.855
	对第二条谣言的愤怒	0.840	
	对第三条谣言的愤怒	0.824	
	对第四条谣言的愤怒	0.829	
恐惧情绪	对第一条谣言的恐惧	0.818	0.828
	对第二条谣言的恐惧	0.819	
	对第三条谣言的恐惧	0.815	
	对第四条谣言的恐惧	0.812	

3.2 描述性统计

描述性统计结果显示,在实验对象(N=160)中,男性本科生40人(25.0%),女性本科生40人(25.0%),男性研究生40人(25.0%),女性研究生40人(25.0%)。这说明实验前对参与者性别和学习阶段的控制达到预期。

在年龄方面(M=21.50, SD=2.445, Max=26, Min=18),其中18-20岁的有63人(39.4%),21-23岁的有51人(31.9%),24-26岁的有46人(28.7%)。

在接受刺激处理后,被试对四则内容的平均愤怒水平(M=3.49, SD=1.387, Max=4.75, Min=1.50)。根据5点李克特量表的定义,实验对象总体偏向于承认对谣言内容感到愤怒。

在接受刺激处理后,被试对四则内容的平均恐惧水平(M=3.51, SD=1.308, Max=4.50, Min=1.75)。根据5级李克特量表的定义,实验对象总体偏向于承认对谣言内容感到恐惧。

在接受刺激处理后,被试的转发总数(M=2.43, SD=1.362, Max=4, Min=0)。其中完全不转发的有21人(13.1%),仅转发一则内容的有23人(14.4%),转发两则内容的有25人(15.6%),转发三则内容的

有49人(30.6%),全部都转发的有42人(26.3%)。

3.3 假设 H1 与 H2 的验证

独立样本 T 检验用于比较两组独立样本的均值,适用于衡量两组样本在连续型变量上的差异是否具有统计学意义。当有实验组和对照组时,独立样本 T 检验可以比较两组的平均水平,从而评估实验处理的效果。因此,本研究通过独立样本 T 检验来分析包含评论组和不包含评论组在愤怒情绪、恐惧情绪、转发数量上的差异水平(见表2)。

结果表明,在假定方差齐次条件下(F=1.244, p=0.266>0.05),包含评论组被试的愤怒情绪水平显著高于不包含评论组(p=0.003<0.01),效应量介于小效应和中等效应之间(Cohen's d=0.481),假设 H1 成立。

此外,在假定方差齐次条件下(F=2.778, p=0.098>0.05),包含评论组被试的恐惧情绪水平显著高于不包含评论组(p=0.006<0.01),效应量介于小效应和中等效应之间(Cohen's d=0.439),假设 H2 成立。

值得注意的是,在假定方差齐次条件下(F=0.441, p=0.508>0.05),包含评论组被试的转发谣言数量显著高于不包含评论组(p=0.027<0.05),效应量介于小效应和中等效应之间(Cohen's d=0.353)。

表2 独立样本 t 检验 (Independent Sample t-test)

变量	组别	M	SD	t	df	p	Cohen's d
愤怒情绪水平	包含评论组	3.81	1.294	3.039	158	0.003	0.481
	不包含评论组	3.16	1.409				
恐惧情绪水平	包含评论组	3.79	1.198	2.777	158	0.006	0.439
	不包含评论组	3.23	1.359				
转发谣言数量	包含评论组	2.66	1.331	2.233	158	0.027	0.353
	不包含评论组	2.19	1.360				

注: M 为样本均值; SD 为标准差; t 为 t 检验统计量; df 为自由度; p 为显著水平; Cohen's d 为效应量

3.4 假设 H3 与 H4 的验证

偏相关分析通过剔除干扰变量的影响,准确地测量两个主要变量之间的相关性,适用于研究连续型变量,是深入探索变量之间线性关系的基础。因此,本研究采用偏相关分析分别检验用户愤怒情绪、用户恐惧情绪与转发谣言数量之间的关系。

首先将被试的性别、年龄、学习阶段、实验分组作为协变量进行控制。结果显示,用户愤怒情绪与谣言转发数量之间存在显著正向关系(r=0.646, p<0.001, df=154),假设 H3 成立。此外,用户恐惧情绪与谣言转发数量之间也存在显著正向关系(r=0.653, p<0.001,

df=154),假设 H4 成立。以上分析结果满足进行中介效应分析的前提条件。

3.5 假设 H5 与 H6 的验证

Process 插件由 Andrew F. Hayes 开发,适用于调节效应、中介效应、调节-中介效应的分析。Model 4 是最常用的模型之一,主要用于研究单个中介变量如何在自变量和因变量之间发挥作用,它不仅能分析自变量通过中介变量影响因变量的间接效应,还可检验自变量对因变量是否存在直接效应。因此,本研究采用 Process 插件 Model 4,检验在社交机器人评论内容对用户谣言转发行为的影响路径上,愤怒情绪是否存在

中介作用。

首先将被试的性别、年龄、学习阶段、恐惧情绪作为协变量进行控制。Model 4 重复采样 5000 次的中介效应检验结果表明（见表 3），浏览评论内容可以显著正向预测被试的愤怒情绪水平（ $p=0.002<0.01$ ， $R^2=0.252$ ）。当浏览评论内容与愤怒情绪水平同时进入回归方程时（ $R^2=0.566$ ），浏览评论内容并不能显著预测被试的谣言转发数量（ $p=0.711>0.05$ ），但此时愤怒情绪水平可以显著正向预测被试的谣言转发数量（ $p<0.001$ ）。

进一步分析结果显示（见表 4），中介效应的 Bootstrap 95% CI 上下限区间不包含 0，表明中介效应显著。直接效应的 Bootstrap 95% CI 上下限区间包含 0，表明直接效应不显著。结合上述，可以判定愤怒情绪水平在浏览评论内容对被试转发谣言数量影响的路径上

具有中介作用，假设 H5 成立。

此外，将被试的性别、年龄、学习阶段、愤怒情绪作为协变量进行控制。Model 4 重复采样 5000 次的中介效应检验结果表明（见表 5），浏览评论内容可以显著正向预测被试的恐惧情绪水平（ $p=0.004<0.01$ ， $R^2=0.248$ ）。当浏览评论内容与恐惧情绪水平同时进入回归方程时（ $R^2=0.564$ ），浏览评论内容并不能显著预测被试的谣言转发数量（ $p=0.535>0.05$ ），但此时恐惧情绪水平可以显著正向预测被试的谣言转发数量（ $p<0.001$ ）。

进一步分析结果显示（见表 6），中介效应的 Bootstrap 95% CI 上下限区间不包含 0，表明中介效应显著。直接效应的 Bootstrap 95% CI 上下限区间包含 0，表明直接效应不显著。结合上述，可以判定恐惧情绪水平在浏览评论内容对被试转发谣言数量影响的路径上具有中介作用，假设 H6 成立。

表 3 多重线性回归（Multiple Linear Regression）

	谣言转发数量			谣言转发数量			愤怒情绪水平		
	Beta	t	p	Beta	t	p	Beta	t	p
浏览评论内容	0.056	0.371	0.711	0.445	2.373	0.019	0.629	3.258	0.002
愤怒情绪水平	0.619	10.233	<0.001						
R^2		0.566			0.269			0.252	
F		33.256			11.333			10.383	

注：Beta 为标准化回归系数；t 为 t 检验统计量；p 为显著水平； R^2 为决定系数；F 为方差检验统计量

表 4 中介效应（Mediating Effect）

	效应系数	平均标准误	95%区间下限	95%区间上限	效应占比
直接效应	0.056	0.150	-0.241	0.352	12.6%
间接效应	0.389	0.128	0.145	0.647	87.4%
总效应	0.445	0.188	0.075	0.815	100.0%

表 5 多重线性回归（Multiple Linear Regression）

	谣言转发数量			谣言转发数量			恐惧情绪水平		
	Beta	t	p	Beta	t	p	Beta	t	p
浏览评论内容	0.093	0.622	0.535	0.445	2.373	0.019	0.540	2.958	0.004
恐惧情绪水平	0.652	10.160	<0.001						
R^2		0.564			0.269			0.248	
F		32.919			11.333			10.151	

注：Beta 为标准化回归系数；t 为 t 检验统计量；p 为显著水平； R^2 为决定系数；F 为方差检验统计量

表 6 中介效应（Mediating Effect）

	效应系数	平均标准误	95%区间下限	95%区间上限	效应占比
直接效应	0.093	0.150	-0.202	0.388	20.9%
间接效应	0.352	0.120	0.127	0.594	79.1%
总效应	0.445	0.188	0.075	0.815	100.0%

3.6 假设 H7 的验证

独立样本 t 检验结果表明（见表 7），在假定方差齐次条件下（ $F=1.347$ ， $p=0.248>0.05$ ），包含法律风险提示组被试的谣言转发数量显著低于不包含提示组（ $p=0.032<0.05$ ），效应量介于小效应和中等效应之间（Cohen's $d=0.296$ ），假设 H7 成立。

3.7 假设 H8 与 H9 的验证

Process 插件 Model 1 适用于执行简单调节分析，它通过检验自变量与调节变量的交互作用来确定调节效应是否存在，能够帮助研究者了解一个自变量对因变量的影响是否会随着调节变量的变化而有所不同。因此，本研究采用 Model 1 来检验在愤怒情绪和恐惧情绪对被试转发谣言数量的影响路径上，来自平台的法律风险提示是否存在调节作用（见表 8）。

将被试的性别、年龄、学习阶段、恐惧情绪、评论浏览作为协变量进行控制。Model 1 重复采样 5000 次的调节效应检验结果表明，当风险提示变量进入回归模型后，风险提示与愤怒情绪的交互项无法显著预测被试的谣言转发数量（ $p=0.324>0.05$ ， $R^2=0.569$ ），Bootstrap 95% CI 上下限区间内包含 0，说明调节效应不显著，假设 H8 不成立。

此外，将被试的性别、年龄、学习阶段、愤怒情绪、评论浏览作为协变量进行控制。Model 1 重复采样 5000 次的调节效应检验结果表明，当风险提示变量进入回归模型后，风险提示与恐惧情绪的交互项无法显著预测被试的谣言转发数量（ $p=0.727>0.05$ ， $R^2=0.564$ ），Bootstrap 95% CI 上下限区间内包含 0，说明调节效应不显著，假设 H9 不成立。

表 7 独立样本 t 检验 (Independent Sample t-test)

变量	组别	M	SD	t	df	p	Cohen's d
谣言转发数量	包含风险提示组	2.65	1.296	2.164	158	0.032	0.296
	不包含风险提示组	2.23	1.405				

注：M 为样本均值；SD 为标准差；t 为 t 检验统计量；df 为自由度；p 为显著水平；Cohen's d 为效应量

表 8 调节效应 (Moderating Effect)

	谣言转发数量					R^2	F
	Beta	t	p	95%区间下限	95%区间上限		
愤怒情绪*风险提示	0.104	0.989	0.324	-0.104	0.312	0.569	28.641
恐惧情绪*风险提示	0.039	0.351	0.727	-0.183	0.262	0.564	28.072

注：Beta 为标准化回归系数；t 为 t 检验统计量；p 为显著水平； R^2 为决定系数；F 为方差检验统计量

4 结果与讨论

4.1 社交机器人生成的谣言评论对大学生微信用户的影响

本研究发现，在浏览网络谣言的基础上，社交机器人针对谣言内容发表的情绪化评论会使大学生微信用户产生更强烈的愤怒和恐惧情绪。一方面，谣言本身通过传播未经证实的信息，尤其是与健康、社会安全有关的内容，会让社交平台用户感到焦虑或恐惧。例如在本研究中，关于校园疾病爆发、食品安全、犯罪活动的虚假信息会导致大学生群体产生不必要的担忧。于此同时，谣言故意捏造的内容以及对事实的夸大与歪曲，通常能煽动起用户的愤怒和仇恨情绪，当人们不断接触到这类信息时，会变得更加敏感和易怒^[18]。另一方面，社交机器人针对谣言内容的评论充斥着大量情绪化表达，在微信群聊这种封闭式群体中很容易形成“回音室

效应”，让某些极端情绪化的观点不断得到重复与加强^[19]。由于用户并不知晓这些评论内容是由社交机器人生成的，因此在群体极化作用下用户的负面情绪也会随之不断增强，进而比单纯浏览谣言内容产生更强烈的愤怒和恐惧。而大学生正好处于心理发展的关键时期，通常在构建自我认同和人生价值观的过程中，这一阶段的认知发展尚未成熟且较易出现情绪波动，因此他们很容易受到网络谣言的影响，尤其是其中包含的负面情绪^[20]。

此外本研究还发现，在浏览社交机器人所发表的情绪化评论后，大学生微信用户会有更高的概率对谣言内容进行分享和转发。理性选择理论认为，人们在做任何决定时是完全理性且自利的，个体会权衡行为的代价和回报，选择能带来最大收益的决策方案。然而现实中人的行为常常偏离这种纯粹的理性，在群体极化

的作用下,理性的思考通常让位于感性的宣泄,使人更容易做出盲从、跟风等非理性的行为而不考虑其后果^[21]。特别是在微信群聊这种非匿名性的公共场合,为了使自己与大多数人的态度和行为保持一致,避免自身陷入被孤立的状态,用户会倾向于模仿其他用户的行为,以此来表现自己合群,并从中获得归属感和认同感。此外,受群体暗示的影响,微信群聊中的用户可能会下意识地作出某些行为,例如对其他用户发表的观点进行附和,这种赞同也许并不是出于内心的真实想法,而是在其他用户暗示下作出的本能回应^[22]。大学生群体对社交媒体的依赖度较高,通常花费大量时间在微信上与同龄人互动,这一方面既增加了他们接触到网络谣言的概率,另一方面也使其容易受到同辈压力的影响,为了让自己保持合群不被孤立而盲目跟随他人,进而提高转发谣言的可能性^[23]。

4.2 愤怒和恐惧情绪对大学生微信用户谣言转发行为的影响

浏览谣言后的愤怒和恐惧水平可以正向预测大学生微信用户的谣言转发概率,且在此过程中愤怒和恐惧情绪具有中介作用。当人处于愤怒或恐惧等强烈情绪中时,理性思考的能力会受到抑制,转而依赖情绪驱动的快速反应。恐惧和愤怒会让人对威胁或危险信息更加敏感,因此更容易相信那些与他们情绪相符的信息,即使这些信息未经验证^[24]。在恐惧状态下,人的大脑专注于寻找潜在威胁,而谣言通常包含夸大或虚构的威胁信息,这正契合了人在恐惧状态下的心理需求,所以更容易被相信。而在愤怒状态下,人则倾向于分享自己所了解的信息,以此获得他人的认同和共情,特别是那些与愤怒源头有关的内容。此外,确认偏见导致人们容易接受那些与自身既有态度一致的信息,而忽视与之相悖的证据^[25]。社交机器人生成的谣言及其评论往往能迎合人们的已有偏见,例如在本研究中,关于食品安全和校园犯罪的谣言更容易激发大学生的愤怒情绪,因为在其相关评论中,情绪宣泄的矛头都指向学校管理层,认为他们的失职是造成这些问题的主要原因,从而诱导被试相信并扩散这些内容,这可能是利用了大学生群体对权威的反抗心理以及对组织管理者的刻板印象。

双系统理论认为,人在决策过程中存在着两种思维方式,一种是快速、自动、直觉化的思维,负责处理日常大多数事物但不经过理性分析思考,另一种则是详细分析的思维模式,会对事物进行多方面的衡量与评估,从而做出理性的判断选择^[26]。在一个封闭式的

群体中,人一旦受到群体情绪的影响,便很容易与他人产生共情,在这种情况下来自他人的信息很有可能被轻易相信并进行二次传播。因为在群体情绪的裹挟下,人会变得高度亢奋,从而抑制自身理性思考和分析判断的思维能力,仅采用快速化的思维模式来处理信息事物^[27]。代入至本研究情境,恐惧和愤怒情绪在微信群聊中尤其容易被感染,因为这些情绪驱动的内容会比中立或理性的内容更能引发大学生用户的共情和互动,而这种互动反馈的循环又能加速谣言的扩散,因为负面情绪让他们更倾向于分享那些能引起共鸣的信息,以此来达到宣泄自身情绪、获得认同和归属感的效果。这一过程也印证了后真相时代下网络舆情危机的特点。

4.3 社交平台法律风险提示对大学生用户谣言转发行为的影响

来自社交平台的法律责任风险提示可以显著降低大学生微信用户的谣言转发概率。有限理性理论认为,人类的理性是有限的,个体在决策时往往无法完全获取或处理所有信息,因而只能作出令自己当下满意的决策,而非长期视角下最优的决策^[28]。该理论还认为人在考虑行为后果时,可能会受到认知、时间和资源的限制,导致决策与完全理性状态下的判断出现偏差。来自平台的法律责任风险提示通过外部介入的方式,强调了转发谣言行为可能要承担的法律后果,帮助用户提高这一因素在决策过程中被考虑的优先级,从而规避了瞬时情绪、时间、注意力资源对决策方案的影响。由于法律责任风险会对用户形成强大的威慑,因此当他们意识到转发谣言可能导致法律后果时,便会变得更加谨慎且不愿轻易转发未经证实的信息。在现实生活中,人们通常倾向于规避影响自身的潜在风险,例如在本研究中,当被试知道转发谣言会导致其个人名誉受损、且可能面临法律惩戒时,他们会倾向于避免这种行为,以免给自己造成不必要的麻烦^[29]。由此可见,法律责任的明确化和用户风险意识的提升,都会促使大学生在面对谣言时更加谨慎,进而降低分享和转发的可能性。

研究结果还显示,法律风险提示并不对用户负面情绪影响转发行为的过程具有调节作用,这说明来自社交平台的法律责任风险提示与大学生微信用户的负面情绪之间不存在明显交互作用,而是作为两个独立的变量所存在,法律风险提示对用户转发行为的影响与愤怒和恐惧情绪的水平无关。这也印证了双系统理论的观点,即当用户被负面情绪裹挟时会倾向于启用

快速化的信息处理模式，而法律风险提示可以通过唤醒用户的风险意识，从而使其切换至理性思维模式以应对后果风险的评估工作^[27]。然而近期有研究指出，双系统并不是两个相互隔离且独立的处理路径，在浏览商品信息到做出购买决策的过程中，两个系统存在着轮流切换的调用机制，且每个系统的处理结果均会对另一方产生影响^[30]。该推论也在另一项通过 fMRI 观察个体决策过程中脑区活跃度变化的研究中得到支持^[31]。因此，用户在接触到社交机器人传播的谣言后，最终的转发与分享行为也可能会受到两种信息处理模式的共同影响，而提前了解转发谣言会面临的法律责任或许有助于提高用户理性思维在决策依据中的权重，使最终行为更偏向于理性。未来需要更多研究考察社交媒体用户面对谣言和错误信息时两种信息处理路径的调用情况，从而为上述理论解释提供更可靠的实证依据。

此外，随着对谣言传播法律后果的了解，用户也会意识到转发谣言的行为可能会遭到他人的批评或抵制，在这种舆论压力下他们会进一步权衡转发的利弊，并加强对自身行为的约束。这一过程通常伴随着用户媒体素养的提升，为了规避法律责任风险，用户倾向于主动学习如何辨别信息的真实性，从而避免不负责任的内容转发，特别是对于在校大学生这一具有较强学习能力的群体^[32]。但是从长期视角来看，关于社交平台法律风险提示是否有助于提升用户媒体素养这一问题还有待进一步考察。

5 研究建议

在理论层面上，该研究通过“知-信-行”行为模式考察了社交机器人生成的谣言及其评论内容在大学生微信用户群聊中的传播机制与影响因素，提示了在生成式人工智能技术背景下谣言传播的规律与特点，并指出大学生群体在面对社交机器人传播的谣言内容时普遍存在群体极化、情绪传染、盲目从众等现象。因此后续研究需要对这些现象及应对策略进行深入探讨，并积极关注技术革新所带来的各种社会风险。

在现实层面上，研究发现针对谣言的情绪化评论内容能影响到其他用户的转发行为。这启示法律法规与平台政策制定者应当重视用户评论在谣言传播过程中的助推作用，因此在针对社交媒体谣言的传播行为制定惩治措施时，不仅要包含内容的制造者和转发分享者，还需把参与评论者囊括在内，尤其是对于那些热衷于发表极端情绪化评论的用户。

此外，社交媒体平台需要不断优化识别算法，通过

分析内容特征和用户行为准确辨别出由社交机器人操纵的帐号，并对这些帐号进行重点监控，提防他们传播谣言进而引发舆情危机的风险。对于普通用户，社交媒体平台可在其评论和转发前弹出相关的法律责任风险提示，类似于银行用户汇款前的转帐风险提示，从而对用户非理性的评论和转发起到警示作用。

最后，对于社会阅历较为浅薄且技术接受能力相对较强的大学生群体，学校应当开设相应的课程来帮助他们提高媒介和网络素养，以此应对人工智能技术普及所带来的潜在风险和挑战。学校一方面需要教导大学生如何辨别社交媒体上信息内容的真伪，另一方面也要帮助他们树立起互联网法律意识，让其学会对自己在网络上的言论和行为负责。

6 研究局限性

本研究尚存在一些局限性。首先，由于研究样本都来自广州大学城且均通过线上招募，因此实验对象的选择无法保证完全随机，需要后续研究进一步验证结论的普适性和可推广性。此外，研究采用自我汇报方法对实验处理效果进行评估，但该方法容易受到社会期望偏差、回忆偏差、自我认知局限、反应倾向、主观解释偏差等因素的影响，这可能会对实验数据的准确性和可靠性造成一定干扰。最后，由于社交媒体使用情境的复杂性，研究设计的 GUI 界面无法完全模拟出大学生微信群聊中的实际情境，例如用户之间的互动和反馈等，这些未被考察的因素可能会对研究变量的作用路径造成某些影响。

参考文献

- [1] 张志勇,荆军昌,李斐,等.人工智能视角下的在线社交网络虚假信息检测、传播与控制研究综述[J].计算机学报,2021,44(11):2261-2282.
- [2] 杨泉泉,王唯滢.社交机器人参与网络伪科学认知传播的机理研究[J].青年记者, 2024(2):23-30.
- [3] 郑晨予,范红.从社会传染到社会扩散: 社交机器人的社会扩散传播机制研究[J].新闻界,2020,(03):51-62.
- [4] 李媛媛,袁玉林,随力瑞.基于 KAP 理论的大学生网络安全教育研究[J].中国安全科学学报, 2024, 34(5):1-8.
- [5] 徐锦江.流言:概念辨析,传播特性及其治理环境研究[J].新闻大学, 2023(8):31-43.
- [6] 顾金喜.“微时代”网络谣言的传播机制研究——一种基于典型案例的分析[J].浙江大学学报(人文社会科学

- 版),2017,47(03):93-103.
- [7] 冉茂洁.基于行为差异和外部干预的谣言传播模型研究与仿真分析[D].重庆理工大学,2019.
- [8] Zhao N, Li Y, Zhang J. A review of influencing factors and motivation mechanisms of rumor spreading[J]. *Journal of Psychological Science*, 2013, 36(4):965-970.
- [9] Nicolas P, Dominik B, Stefan F. Emotions in online rumor diffusion[J]. *EPJ Data Science*, 2021, 10(1):51.
- [10] 赖胜强,唐雪梅.信息情绪性对网络谣言传播的影响研究[J]. *情报杂志*, 2016, 35(01):116-121.
- [11] 周书环,杨潇坤.新冠肺炎疫情下社交媒体情绪传播及其影响研究——基于新浪微博文本数据的实证分析[J]. *新闻大学*, 2021, (08):92-106+120-121.
- [12] 刘晓燕,张乐乐.后真相时代社交媒体情绪极化研究[J]. *青年记者*, 2024, (07):84-88.
- [13] 杨颖.社交媒体舆情信息热度和情感强度对传播意愿的影响——情绪与感知可信度的双重中介机制[J]. *心理学进展*, 2022, 12(7):2424-2432.
- [14] 丁绪武,吴忠,夏志杰.社会媒体中情绪因素对用户转发行为影响的实证研究——以新浪微博为例[J]. *现代情报*, 2014, 34(11):147-155.
- [15] 唐雪梅,赖胜强.公共卫生安全事件中网络健康谣言的转发研究——感知风险与信息可信度的交互效应[J]. *情报杂志*, 2021(9):101-107.
- [16] 王钰昱.添加警告对不同类型在线健康谣言分享意愿的影响研究[J]. *应用数学进展*, 2022, 11(9):6288-6297.
- [17] LYU X, LIU X, YANG T, et al. Intervention strategies for health rumors: An overview based on the information lifecycle theory[J]. *Advances in Psychological Science*, 2024, 32(4):664-676.
- [18] 杨慧琼,阙耀龙,麻小影.谣言传播背后的情绪动力机制——基于食品安全谣言的调查研究[J]. *东南传播*, 2019, (08):94-100.
- [19] 何飞,杨森.情绪性谣言传播机制及演化机理研究——以新冠肺炎疫情为例[J]. *当代传播*, 2022(1):90-93.
- [20] 张益萌,王馨,董子泽,等.突发公共卫生事件网络谣言大学生群体认知差异研究[J]. *中国卫生法制*, 2024, 32(2):44-49.
- [21] Hechter M, Kanazawa S. Sociological rational choice theory[J]. *Annual review of sociology*, 1997, 23(1):191-214.
- [22] 马相龙,卯会.警惕自媒体中“媒介审判”下的群体暗示行为——以“红黄蓝”事件为例[J]. *新媒体研究*, 2018, 4(09): 121-122.
- [23] Yelishala A. The impact of social media on peer pressure in adolescents[J]. *Youth Medical Journal*, 2022.
- [24] 曹正,何佳,张华春.网络非理性舆论对大学生思想行为影响研究[J]. *西南石油大学学报(社会科学版)*, 2023, 25(04): 101-110.
- [25] Fernandes M R. Confirmation bias in social networks[J]. *Mathematical Social Sciences*, 2023, 123:59-76.
- [26] Evans J S B T, Stanovich K E. Dual-process theories of higher cognition: Advancing the debate[J]. *Perspectives on psychological science*, 2013, 8(3):223-241.
- [27] 袁媛,苏红云.从群体极化理论探索后真相时代下的舆论引导[J]. *南方传媒研究*, 2023, (06):61-67.
- [28] Gigerenzer G. What is bounded rationality?[M]// *Routledge handbook of bounded rationality*. Routledge, 2020:55-69.
- [29] Huber O, Huber O W, Bär A S. Framing of decisions: Effect on active and passive risk avoidance[J]. *Journal of Behavioral Decision Making*, 2014, 27(5):444-453.
- [30] Li Y, Xu Z, Teo T. Harnessing dual process theory for online product evaluation based on user-generated content[J]. *Industrial Management & Data Systems*, 2025.
- [31] Gronchi, G., Gavazzi, G., Viggiano, M. P., & Giovannelli, F. (2024). Dual-Process Theory of Thought and Inhibitory Control: An ALE Meta-Analysis. *Brain Sciences*, 14(1):101.
- [32] 黄华.辨识与防范:青年群体应对网络谣言圈群化传播风险的素养与能力研究[J]. *新媒体研究*, 2024, 10(7):80-85.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS