交互性预先干预对社交平台用户虚假信息感知和行为的影响*

杨雪艳　夏志杰
（上海工程技术大学管理学院，上海 201620）

摘要：通过将基于逻辑接种的预先干预措施与会话式人工智能相结合，验证智能系统驱动的交互性预先干预对社交平台用户虚假信息感知及行为（点赞和转发）的影响，为社交平台打击虚假信息提供新思路和方法。基于S-O-R理论构建研究框架模型，设计对照实验，采用问卷调查法收集数据，并通过单因素方差分析和中介效应分析探讨干预措施对用户虚假信息感知和行为的差异性影响。研究结果表明：该干预措施显著减少了虚假信息的感知可信度、感知准确性、感知真实性以及用户的点赞和转发行为；同时，感知真实性和感知准确性在干预措施与用户行为之间起到一定的中介作用。研究建议，社交平台可通过预先干预措施减少虚假信息的负面影响，并在制定干预策略时综合考虑多重因素。

关键词：社交平台；虚假信息；预先干预；个体感知；信息行为；S-O-R框架模型
中图分类号：G206　DOI：10.3772/j.issn.1673-2286.2024.11.005
引文格式：杨雪艳，夏志杰. 交互性预先干预对社交平台用户虚假信息感知和行为的影响[J]. 数字图书馆论坛，2024，20（11）：43-51.

随着信息技术的发展和人工智能生成内容的广泛应用，信息的生成和传播过程变得愈加复杂和动态化，使得打击虚假信息变得越来越困难[1]。社交平台的普及极大地加剧了这一情况，使得虚假信息能够迅速传播并广泛分享给不同公众[2]。心理预防作为一种预防性治疗措施，因可以有效弥补事实核查和揭露等虚假信息干预中存在的一些不足，得到人们的广泛关注[3]。人工智能和大语言模型的技术进步为人工智能与多种干预措施的结合提供了新可能，在提升心理预防效果和打击虚假信息方面成为热门话题[4]。

虚假信息的干预措施主要包括事实核查和揭露，其效果受到辟谣信息质量、公众已有观念及意识形态等多方面因素的影响[5]。“虚幻真相效应”表明，纠正虚假信息可能适得其反，因为人们更容易相信重复的信息（即虚假信息）而非真实信息；而“持续影响效应”则表明，即使采取干预措施，人们也可能对最初的虚假信息产生信任[6]。为应对这些挑战，研究人员开始探索针对虚假信息的心理预防策略。通过预先的警告和驳斥，在公众接触虚假信息或其具有“传染性”之前，培养公众对未来虚假信息的认知抵抗力[5]。近年来，这一领域已从关注培养个体抵御特定虚假信息的能力，逐步转向关注虚假信息背后的潜在策略，实现从“窄谱”（基于事实）的接种向“广谱”（基于逻辑）的接种转变。与事实接种相比，逻辑接种显著提高了干预措施的可扩展性，使公众能够对一系列虚假信息形成更强的免疫力[7]。此外，会话式人工智能可以改善用户交互体验，吸引用户注意力，进而影响用户认知、态度和行为[8]。会话式人工智能在虚假信息干预方面具有潜在的积极

收稿日期：2024-07-05
*本研究得到国家社会科学基金一般项目“大数据支持下网络谣言智慧治理机制及运行策略研究”（编号：21BGL243）、上海市哲学社会科学规划一般项目“大数据时代伪健康信息传播特征及多主体协同干预研究”（编号：2020BGL005）资助。

影响，利用研究对象与人工智能之间实时、个性化的交互，能够有效降低他们对虚假信息的信任[4]。

鉴于此，本文尝试将基于逻辑接种的预先干预措施和会话式人工智能相结合，以刺激-机体-反应（Stimulus-Organism-Response，S-O-R）理论为理论框架，验证具有交互性质的基于逻辑接种的预先干预措施是否会对公众（尤其是社交平台用户）虚假信息感知和行为产生影响，旨在为打击虚假信息提供一种新的思路，减少虚假信息的负面影响。

1 文献回顾

1.1 虚假信息应对策略

虚假信息应对策略研究深受各界关注。从信息传播的时间维度来看，虚假信息的应对策略有3种：传播前的预防策略、传播中的阻断策略和传播后的纠正策略。以技术为主导的事中阻断和以事实核查与反驳为核心的事后纠正是当下流行的虚假信息应对方式[9]。Kimura等[10]提出一种自然贪婪算法并将其应用于阻塞虚假信息传播。考虑突发事件信息传播生命周期的TPR方法[11]、两阶段贪婪算法[12]、基于竞争性独立级联模型提出的CDRBM算法[13]等均被证实能够有效阻断虚假信息的传播。现有研究中，虚假信息的检测通常被视为一个分类问题，研究依靠自然语言处理、机器学习、图论等多种方式来有效区分虚假信息和真实信息[14]。虚假信息阻断策略主要通过阻塞虚假信息传播路径、控制虚假信息传播节点、提高虚假信息的传播阻力来实现，但阻断虚假信息传播的过程也会影响真实信息的传播效率以及用户对社交平台的使用体验。

对于已经受到虚假信息污染的信息环境，则需要及时纠正虚假信息并消除虚假信息的影响。微信官方辟谣账号“谣言过滤器”、“科普中国”等微信公众号以及中国互联网联合辟谣平台通过定期向用户推送已辟谣的虚假信息或谣言，帮助用户了解并识别当前遇到的虚假信息。然而，事实核查信息的传播广度远小于虚假信息的传播广度，加之用户态度和认知能力等因素影响，事后纠正存在的“回音室效应”“持续影响效应”“虚幻真相效应”等认知偏差难以完全消除[1]，因此虚假信息传播后的纠正策略不仅存在滞后性，效果也十分有限。

传播前的预防策略对真实信息传播效率影响较小，同时能减少认知偏差并降低治理成本，因此得到国内外研究者的关注。van der Linden等[15]提及可以通过针对虚假信息的“接种”或“预先干预措施”来应对虚假信息问题。

1.2 接种理论和预先干预措施

接种理论最先在气候变化[16]、政治[17]等相关虚假信息问题上进行了验证。虚假信息预先干预措施以接种理论为框架，通过让人们接受减弱剂量的特定的论点来建立心理抵抗力，以防范未来虚假信息的影响[18]。接种理论作用机制包括通过事实证据或逻辑技巧对攻击进行先发制人的反驳，以此形成两种干预措施[19-20]：事实接种，即基于事实证据对虚假信息中的论点进行预先驳斥；逻辑接种，即专注于构成虚假信息的操纵技巧和逻辑谬误，而非特定情境的个别例子。现有研究大多集中于事实接种，随着心理预防领域的发展，越来越多的人关注逻辑接种领域，这也促进心理预防从“基于事实”到“基于技术”的转变。

社交平台虚假信息常用的操纵技巧有情绪化语言、不连贯（或没有条理）、假二分法（或虚假两难的逻辑谬误）、替罪羊、人身攻击等[3]。研究表明，向社交平台用户介绍新冠疫苗失真环境中常见的信息操纵策略[19]，或将构成虚假信息的操纵技术融入小游戏[7]，能够有效提升用户对虚假信息的心理防御能力，同时降低虚假信息的感知可信度并减少虚假信息分享行为。传统接种研究大多构建对特定说服性攻击的抵御能力，而关注支撑虚假信息的操纵技术和策略能显著提高接种干预在社交平台上的可扩展性[7]。通过将包含逻辑接种的接种视频作为广告投放于YouTube等社交平台，证实了干预措施在现实环境中的有效性[3]。实验表明，无论是实验室环境还是经常接触虚假信息的现实环境，人们均可以通过观看简短的接种视频来提高识别虚假信息的能力；而且这些干预措施对具有不同意识形态背景和认知风格的人均是有效的，可以很容易地大规模实施。

1.3 交互性预先干预

社交互动是公众获取信息、交流观点的重要途径，

能有效降低认知风险，帮助甄别信息[21]。人们可以通过与生成式人工智能（如ChatGPT）互动来获得社交互动的满足[22]。会话式人工智能是人工智能领域的新兴方向，可以通过询问用户来启动会话以及指导会话流程，并显著提升人机交互能力[23]。例如，微软和谷歌将会话式问答添加到搜索引擎中，以此改善用户的互动体验。交互性是人机交互的基础，在重塑人机交互方面发挥着关键作用。根据信息传播的三大要素（来源、媒介和信息），可将交互性划分为媒介交互性、来源交互性、信息交互性[8]。在数字媒体领域，互动变得无所不在，交互性可以提供线索和启示，吸引用户注意力，并影响用户的认知、态度和行为意图[24]。

基于上述讨论，本文所述交互性预先干预指的是：将预先干预的相关文本内容与会话式人工智能系统相结合，嵌入社交平台，以此提高预先干预措施的交互性。交互性预先干预的独特之处在于提高参与者的关注度以及调节参与者认知和思维。有研究表明，数字互动干预是促进老年人心理健康的有效方法之一，在缓解抑郁和焦虑方面有一定应用[25]。为用户提供交互性环境可以增强自我效能感，进而增加现实生活中行为改变的可能性[26]。Hurling等[27]开发的自动对话模块促进了健康相关行为信念的形成。总之，具备高度交互性和吸引力的会话式人工智能系统能够增强用户的感知控制，并对行为产生积极的影响。

综上，传播前的预防策略与其他虚假信息应对策略相比具有一定的优势：对真实信息传播效率影响较小，同时能减少认知偏差。而且，基于逻辑的干预方式与基于事实的干预方式相比具有一定的可扩展性，同时交互性可以有效提高用户的参与度，并影响用户的感知、态度和行为等方面。故本研究试图解决以下两个问题：①在基于逻辑接种的预先干预措施中增加交互是否具有良好的效果；②交互性预先干预措施如何影响社交平台用户对虚假信息的感知与行为。利用问卷调查收集数据，基于SPSS软件进行数据分析，将社交平台用户感知维度分为感知可信度、感知准确性、感知真实性，行为维度分为点赞行为和转发行为，从而分析交互性预先干预措施对社交平台用户感知和行为的影响。

2　研究假设

S-O-R模型适用于解释个体在面对环境刺激时内在的感知、态度、信念以及相应的行为反应。周涛等[28]将内外部动机作为刺激因素，选择焦虑情绪和社交网络疲劳两个因素作为机体因素，来研究在线知识社区用户的潜水行为。唐雪梅等[29]则发现焦虑是影响健康谣言转发的中介变量，受众转发行为遵循S-O-R模式框架。基于S-O-R模型构建研究框架模型（见图1）。

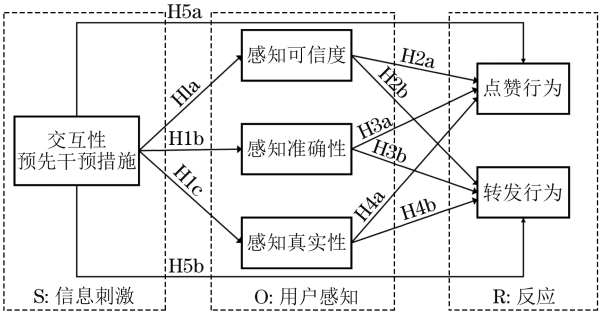

图1　研究框架模型

2.1　交互性预先干预措施对个体虚假信息感知的影响

作为中介变量，用户感知代表个体的心理状态变化，例如情感反应或认知反应等。用户对信息的感知包含多个层面，如感知可信度、感知准确性、感知威胁、来源可信度、感知质量等。对虚假信息的感知不仅包括可信度、准确性，真实性也是重要的指标之一[30]。因此，对于虚假信息的用户感知，从可信度、准确性、真实性3个维度进行考虑。

2.1.1　感知可信度

可信度被定义为用户感知到的信息质量。感知可信度是信息接受者对信息的整体感知判断，常被解构为一个多维概念，如可信任性、专业性等[19]。如何降低虚假信息的可信度以降低公众对科学知识的误解程度是国内外学者关注的热点问题。使用克服公关虚假信息攻击的预先谎言策略可以降低攻击组织的感知可信度，有助于防止虚假信息在社交平台上的传播[31]。Roozenbeek等[32]在干预措施的跨文化研究中也证实，预先干预措施可以有效降低虚假信息的感知可靠性，提高参与者发现虚假信息的能力。现有研究结果表明，事实接种和逻辑接种两类预先干预措施均能够显著降低失真健康信息的感知可信度[19]。因此，提出以下假设。

H1a：交互性预先干预措施可以降低社交平台上虚假信息的感知可信度。

2.1.2 感知准确性

感知准确性在人们判别假新闻时发挥着作用，涉及个体对外界信息的准确解读和理解，是个体对外界信息的感知与实际信息之间的符合程度[33]。当人们愿意进行更多的思考和推理的时候，对虚假信息的感知准确性降低。Sharevski等[34]的研究表明人们预先存在的信念会影响个人对社交平台信息准确性的看法，在消除人们对虚假信息的信念方面，覆盖所有内容、要求用户点击后才能查看内容的插页式封面的警告形式，比位于社交平台内容下方、带有警告性质的上下文标签的警告形式更有效果。Clayton等[35]发现，在文章中添加“评级错误”标签比添加“有争议”标签更能降低感知准确性，且不会对未标记信息的感知准确性产生影响。基于上述研究论述，提出如下假设。

H1b：交互性预先干预措施可以降低社交平台上虚假信息的感知准确性。

2.1.3 感知真实性

基于线索效用理论的研究结果表明，分享者信任度、分享者观点、附加提示等因素均会影响深度伪造信息的感知真实性和用户的互动行为（点赞、评论、转发），并且这些因素之间还存在显著的交互效应[30]。同时也有实验表明，感知到的夸张在多大程度上可以解释用户感知到的评论真实性取决于酒店的类别和相关评论的极性，而更深层次的影响因素则与用户的确认偏误等认知偏差有关[36]。用户对信息真实性的感知受到认知偏差的影响，基于逻辑接种的预先干预措施可以影响人们的认知方式。因此，提出以下假设。

H1c：交互性预先干预措施可以降低社交平台上虚假信息的感知真实性。

2.2 交互性预先干预措施对个体虚假信息传播行为的影响

反应是代表个体态度或行为的结果变量，通常表现为对某一事物的趋近或规避。社交平台虚假信息具有数量多、传播广等特点，信息参与行为是公众接触虚假信息的后续行为，个体对虚假信息的传播行为（如点赞、转发、评论等）会进一步促进虚假信息的传播与扩散，加大信息治理的难度[37]。点赞行为是一种由情感驱动的行为，而转发行为会受到认知的影响[38]，故从点赞行为和转发行为两个方面来分析干预措施对社交平台用户行为的影响。感知可信度和感知准确性反映的是信息给用户带来的权威印象，感知可信度会影响用户在社交平台分享信息的决定，感知准确性对其分享决策有至关重要的作用[39]。预先干预能够在虚假信息传播之前赋予个体抗性，降低个体对虚假信息的分享意愿，防止虚假信息在社交平台上的传播。研究表明，游戏作为预先干预的方式之一，不仅能够降低虚假信息的感知可靠性，增强自主识别虚假信息的能力，还能显著降低社交平台用户分享虚假信息的意愿，且这种干预在实施后3个月内依旧有效[40-41]。综上所述，提出以下假设。

H2a：虚假信息的感知可信度影响用户对虚假信息的点赞行为。

H2b：虚假信息的感知可信度影响用户对虚假信息的转发行为。

H3a：虚假信息的感知准确性影响用户对虚假信息的点赞行为。

H3b：虚假信息的感知准确性影响用户对虚假信息的转发行为。

H4a：虚假信息的感知真实性影响用户对虚假信息的点赞行为。

H4b：虚假信息的感知真实性影响用户对虚假信息的转发行为。

H5a：交互性预先干预措施可以减少用户对社交平台上虚假信息的点赞行为。

H5b：交互性预先干预措施可以减少用户对社交平台上虚假信息的转发行为。

3 研究方法

3.1 实验材料设计

通过对照实验和问卷分析来探究具有交互性的基于逻辑接种的预先干预措施是否会对社交平台用户虚假信息感知和行为产生影响。接种内容主要包含虚假

信息常见操纵策略（情绪化语言、不连贯或没有条理、假二分法或虚假两难的逻辑谬误、替罪羊、人身攻击）的具体定义、判定方法及对应例子。实验组在回答问卷之前会浏览接种内容并想象自己正在和会话式人工智能系统聊天，而对照组在回答问卷之前不接受任何的干预。

虚假信息预先干预措施涵盖多个主题，主要集中在气候变化、政策和健康3个方面，且有效性均已得到初步验证[42]。健康相关的虚假信息是网络上常见的虚假信息类型之一，是网络治理的突出议题[21]。因此，参考科学辟谣平台、中国互联网联合辟谣平台等相关网站上已被证实为虚假的3则信息，并按照微博内容的发布特点进行修改，设计了“支原体肺炎引发白肺”“吃大剂量维生素C可以预防流感”“适量饮酒有助于血管健康”3则健康类虚假信息。对照组和实验组均阅读相同的虚假信息材料，3则材料在字数上基本保持一致，在语言表达上使用了假设情境中提及的虚假信息操纵技巧。同时，为了避免3则虚假信息给被试人员带来一定的影响，在问卷调查结束时通过信息或口头的方式告知实验对象3则虚假信息的真伪以及权威人士对3则信息的辟谣内容。

3.2　实验分组

实验分为对照组和实验组，采用随机对照实验法，即参与者和实验者均不知晓实验干预与对照组的分配，以消除主观偏见和心理暗示对结果的影响，从而获得更客观、可靠的结论。在问卷调查前，询问被试者的微博使用情况，若被试者近期未使用微博，则不进行后续实验。

3.3　变量测量

变量测量分为参与者个体感知和行为两个维度以及人口统计变量（性别、年龄、受教育程度）测量。个体感知分为3个维度：感知可信度、感知准确性、感知真实性；个体行为分为两个维度：点赞行为、转发行为。感知和行为的测量均通过问卷进行，在每则健康类虚假信息下设置“请根据自我认知对该微博截图内容的可信度/准确性/真实性进行评分”“您有多大可能对该信息点赞”“您有多大可能在网上分享该信息”等问题。感知和行为方面问题均采用李克特5级量表测量。同时出于质量控制的目的，在问卷中设计了题为“出于质量控制的目的，请在这里选择‘略微不同意’”的注意力测试题目。

4　数据分析

4.1　描述性统计分析

问卷主要通过问卷星平台进行分发，剔除没有通过注意力测试以及存在少填等明显错误的问卷，最终有效问卷共115份。有效问卷中共有50人自动选择为实验组，其余的65人则为对照组。

由人口统计结果可知，被试者大多为年轻人（19~25岁占83.48%，26~35岁占13.04%），男女比例约为4∶6，比较符合微博的用户画像（《2023微博年轻用户发展报告》指出：微博中16~22岁的活跃用户规模超过1.3亿，男女比例为44∶56），因此本问卷调查样本在一定程度上可以代表微博平台的用户群体。为了保证实验数据的有效性，对问卷数据的信效度进行分析。结果显示，感知和行为维度量表的Cronbach’s α值分别为0.900和0.910，总量表的Cronbach’s α值为0.937（均大于0.8），且两部分内容在删除任意项后的Cronbach’s α值均小于删除之前，故可以认为量表具有较高的信度。KMO值为0.856（$P<0.001$），大于0.8且显著；利用Amos 28软件进行验证性因子分析，结果显示因子载荷为0.520~0.825，并通过计算得到平均提取方差为0.504（大于0.5）、组合信度为0.937（大于0.7），问卷通过效度检验。

4.2　单因素方差分析

利用SPSS软件进行方差分析，对实验组和对照组间的随机化分组效果进行检验。对照组与实验组各协变量单因素方差分析（见表1）显示，各协变量P值均大于0.05，不存在显著系统性差异，因此分组较为随机。

预先干预行为对感知和行为的影响如表2所示。结果表明，对照组与实验组的感知可信度有显著差异（F=8.123，df1=1，df2=113，P=0.005），实验组的均值（M=2.963，σ=0.785）要明显低于对照组（M=3.405，

σ=0.853），假设H1a得到支持。同样，在感知准确性和感知真实性方面对照组和实验组之间也存在显著差异，感知准确性和感知真实性实验组均值要明显低于对照组，因此假设H1b与假设H1c均得到支持。

通过方差分析探讨了交互性预先干预措施对微博平台用户点赞和转发行为的影响。结果表明，对照组和实验组的点赞行为和转发行为存在着一定的差异（点赞行为：F=4.597，P=0.034；转发行为：F=6.563，P=0.012），实验组的均值明显低于对照组，故假设H5a和假设H5b得到验证。

表1　对照组与实验组各协变量单因素方差分析结果

变量	组别	均值M	标准偏差σ	F	P
年龄	对照组	2.215	0.515	3.536	0.063
	实验组	2.060	0.314		
性别	对照组	0.600	0.494	0.429	0.514
	实验组	0.660	0.479		
受教育程度	对照组	3.908	0.723	0.008	0.929
	实验组	3.920	0.752		

表2　感知和行为维度单因素方差分析结果

变量	组别	均值M	标准偏差σ	F	P
感知可信度	对照组	3.405	0.853	8.123	0.005
	实验组	2.963	0.785		
感知准确性	对照组	3.303	0.814	7.553	0.007
	实验组	2.880	0.821		
感知真实性	对照组	3.308	0.915	7.821	0.006
	实验组	2.847	0.823		
点赞行为	对照组	2.780	1.123	4.597	0.034
	实验组	2.353	0.963		
转发行为	对照组	2.564	1.174	6.563	0.012
	实验组	2.053	0.889		

4.3　中介效应分析

为了进一步了解自变量（有无干预措施）影响社交平台用户的点赞和转发行为（因变量）的具体路径，运用SPSS软件的PROCESS Macro Model 4（简单中介）来测量感知可信度、感知准确性、感知真实性的中介效应。

点赞行为的中介效应分析结果如表3所示。直接效应分析显示，交互性预先干预措施与用户点赞行为之间的直接联系并不显著，因为两者直接效应的95%置信区间为[–0.295 6，0.282 8]，包含了0。在间接效应分析中，总的间接效应是显著的，95%置信区间为[–0.703 7，–0.121 7]，不包含0。进一步地，分析中介变量的效应时发现，感知准确性的中介作用显著，其95%置信区间为[–0.359 3，–0.027 3]，不包含0。相反，感知可信度和感知真实性的中介作用不显著，因为它们的95%置信区间包含了0。基于上述分析，感知准确性在干预措施与用户点赞行为之间具有完全中介效应。这表明，交互性预先干预措施主要通过改变信息的感知准确性来影响用户的点赞行为，交互性预先干预措施通过感知准确性的中介作用对用户的点赞行为产生了显著的负向影响。

转发行为的中介效应分析结果如表4所示。直接效应分析显示，交互性预先干预措施与用户转发行为之间的直接联系并不显著，两者直接效应的95%置信区间为[–0.391 9，0.193 9]，包含了0。在间接效应分析中，总的间接效应是显著的，因为95%置信区间为[–0.712 4，–0.127 2]，不包含0。进一步分析发现，感知真实性对干预措施与转发行为之间的关系起到了显著的中介作用，其95%置信区间为[–0.470 9，–0.018 8]，不包含0。

表3　点赞行为的中介效应分析结果

类别		效应值	标准误	t	P	95%置信区间	
						下限	上限
直接影响		–0.006 4	0.145 9	–0.043 9	0.965 1	–0.295 6	0.282 8
间接影响	总体	–0.419 7	0.146 0			–0.703 7	–0.121 7
	感知可信度	–0.147 0	0.109 1			–0.407 9	0.018 2
	感知准确性	–0.178 5	0.087 1			–0.359 3	–0.027 3
	感知真实性	–0.094 2	0.111 5			–0.338 9	0.124 3

同样，感知准确性也起到一定的中介作用，95%置信区间为[–0.351 5，–0.003 2]，接近0，但不包含0，表明该效应显著但效应较小。相反，感知可信度作为中介变量时，95%置信区间为[–0.267 3，0.107 7]，包含了0，中介效应不显著。基于上述分析，交互性预先干预措施对用户转发行为的负向影响主要通过感知真实性和感知准

表4　转发行为的中介效应分析结果

类别		效应值	标准误	*t*	*P*	95%置信区间	
						下限	上限
直接影响		−0.099 0	0.147 8	−0.669 9	0.504 3	−0.391 9	0.193 9
间接影响	总体	−0.411 8	0.148 9			−0.712 4	−0.127 2
	感知可信度	−0.050 4	0.093 1			−0.267 3	0.107 7
	感知准确性	−0.153 4	0.089 9			−0.351 5	−0.003 2
	感知真实性	−0.208 0	0.116 9			−0.470 9	−0.018 8

确性的中介作用来体现。

通过上述中介效应分析可以得出假设H3a、H3b与H4b均得到支持，而假设H2a、H2b、H4a并未得到有效的证实。

5　结论

研究发现，具有交互性的基于逻辑接种的预先干预措施能够有效降低虚假信息的感知准确性、感知可信度和感知真实性，验证了其在增强用户虚假信息感知方面的有效性。这种方法在未显著影响用户对真实信息的辨识能力的同时，还对用户行为产生了积极影响，具体表现为减少了用户对虚假信息的点赞与转发行为。中介效应分析进一步揭示了点赞行为与转发行为的内在逻辑差异：用户对信息准确性的判断主要影响点赞行为，而感知真实性和感知准确性共同影响转发行为。这一发现表明不同心理因素在用户行为中的差异性作用。基于此，建议将预先干预措施融入会话式人工智能系统并嵌入社交平台，以增强干预措施交互性、普适性和推广性。通过个性化设计，不同用户群体可以获得定制化的干预体验。此外，该系统还能借助大数据与算法，为用户提供更加精准的虚假信息防范策略。

本研究不仅深化了对预先干预措施作用机制的理解，还为社交平台上的虚假信息治理提供了切实可行的实践基础。通过预先干预措施的实施，可以有效提升用户的认知，充分发挥用户在虚假信息治理中的主体作用，实现虚假信息治理的科学化和系统化。

未来研究可以从以下几个方面进一步探索。首先，在现实情境中进行相应研究，并通过访谈、观察、行为追踪和生理数据测量等多种数据收集方法和更加专业的测量工具来提高数据的可靠性和普适性。其次，结合社会认知理论、精细加工可能性模型等来丰富分析框架，深入探讨预先干预措施更加具体的干预路径，全面理解其内在机制，进一步优化干预措施。最后，增加实验材料的主题类型，以及使用语言风格更多样的干预材料，优化逻辑接种干预的设计。

参考文献

[1] 宋士杰，赵宇翔，朱庆华. 社交媒体中失真健康信息的传播、识别与纠偏研究[J]. 情报杂志，2023，42（6）：162-169.

[2] NOMAN M，GURGUN S，PHALP K，et al. Designing social media to foster user engagement in challenging misinformation：a cross-cultural comparison between the UK and Arab countries[J]. Humanities and Social Sciences Communications，2024，11（1）：1045.

[3] ROOZENBEEK J，VAN DER LINDEN S，GOLDBERG B，et al. Psychological inoculation improves resilience against misinformation on social media[J]. Science Advances，2022，8（34）：eabo6254.

[4] COSTELLO T H，PENNYCOOK G，RAND D G. Durably reducing conspiracy beliefs through dialogues with AI[J]. Science，2024，385（6714）：eadq1814.

[5] VAN LINDEN S. Misinformation：susceptibility，spread，and interventions to immunize the public[J]. Nature Medicine，2022，28（3）：460-467.

[6] 李松涛，张卫东，陈希鹏. 辟谣信息对社交媒体用户错误信息持续影响效应（CIE）的影响：基于知识修正理论[J]. 现代情报，2024，44（5）：123-139.

[7] ROOZENBEEK J，VAN DER LINDEN S. Fake news game confers psychological resistance against online misinformation[J]. Palgrave Communications，2019，5（1）：65.

[8] SUNDAR S S，XU Q，BELLUR S. Designing interactivity in media interfaces：a communications perspective[C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York：ACM Press，2010：2247-2256.

[9] 李新月，练靖雯，刘周颖，等. 健康信息治理新路径：失真健康信息预先干预的实践案例分析与经验启示[J]. 图书馆杂志，2023，42（12）：120-132，100.

[10] KIMURA M，SAITO K，MOTODA H. Blocking links to minimize contamination spread in a social network[J]. ACM Transactions on Knowledge Discovery from Data，2009，3（2）：1-23.

[11] 陈思菁，李纲，毛进，等. 突发事件信息传播网络中的关键节点动态识别研究[J]. 情报学报，2019，38（2）：178-190.

[12] 倪培昆，朱建明，王国庆. 在线社交网络虚假信息交互量最小化的边阻断策略研究[J]. 中国管理科学，2021，29（9）：188-200.

[13] 刘维，杜宁宁，陈岐，等. CIC模型下基于社区检测的谣言抑制最大化方法[J]. 南京大学学报（自然科学），2023，59（2）：282-294.

[14] AÏMEUR E，AMRI S，BRASSARD G. Fake news，disinformation and misinformation in social media：a review[J]. Social Network Analysis and Mining，2023，13（1）：30.

[15] VAN DER LINDEN S，ROOZENBEEK J. Psychological inoculation against fake news[M]//The Psychology of Fake News. London：Routledge，2020：147-169.

[16] SCHMID-PETRI H，BÜRGER M. The effect of misinformation and inoculation：replication of an experiment on the effect of false experts in the context of climate change communication[J]. Public Understanding of Science，2022，31（2）：152-167.

[17] BANAS J A，MILLER G. Inducing resistance to conspiracy theory propaganda：testing inoculation and metainoculation strategies[J]. Human Communication Research，2013，39（2）：184-207.

[18] MCGUIRE W J. Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments[J]. The Journal of Abnormal and Social Psychology，1961，63（2）：326-332.

[19] 李新月，宋士杰，韩文婷，等. 失真健康信息预先干预对个体感知与行为意愿影响的实证研究[J]. 信息资源管理学报，2024，14（1）：44-54.

[20] NEYLAN J，BIDDLESTONE M，ROOZENBEEK J，et al. How to “inoculate” against multimodal misinformation：a conceptual replication of roozenbeek and van der linden（2020）[J]. Scientific Reports，2023，13（1）：18273.

[21] 夏志杰，吕柠. 智能系统支持下社交互动对提升公众信息甄别能力的影响研究[J]. 情报资料工作，2024，45（5）：41-50.

[22] PATARANUTAPORN P，LIU R，FINN E，et al. Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness，empathy and effectiveness[J]. Nature Machine Intelligence，2023，5：1076-1086.

[23] SAKA A B，OYEDELE L O，AKANBI L A，et al. Conversational artificial intelligence in the AEC industry：a review of present status，challenges and opportunities[J]. Advanced Engineering Informatics，2023，55：101869.

[24] KOZAK J，FEL S. How sociodemographic factors relate to trust in artificial intelligence among students in Poland and the United Kingdom[J]. Scientific Reports，2024，14：28776.

[25] SHI X Y，ZHANG J X，WANG H L，et al. The effectiveness of digital interactive intervention on reducing older adults’ depressive and anxiety symptoms：a systematic review and meta-analysis[J]. Gerontology，2024，70（9）：991-1012.

[26] CHEN T，LI X，DUAN Y Q. The effects of cognitive dissonance and self-efficacy on short video discontinuous usage intention[J]. Information Technology & People，2024，37（4）：1514-1539.

[27] HURLING R，FAIRLEY B W，DIAS M B. Internet-based exercise intervention systems：are more interactive designs better? [J]. Psychology & Health，2006，21（6）：757-772.

[28] 周涛，刘佳怡，邓胜利. 基于SOR模型的在线知识社区用户潜水行为研究[J]. 情报杂志，2022，41（7）：160-165，83.

[29] 唐雪梅，赖胜强. 基于焦虑视角的受众网络健康谣言转发意愿研究[J]. 情报杂志，2023，42（11）：135-141.

[30] 刘春年，陈梦秋. 基于线索效用理论的用户虚假信息感知与互动行为研究：面向深度伪造信息[J]. 情报杂志，2023，42（10）：96-104，19.

[31] BOMAN C D. Examining characteristics of prebunking strategies to overcome PR disinformation attacks[J]. Public Relations Review，2021，47（5）：102105.

[32] ROOZENBEEK J，VAN DER LINDEN S. Inoculation theory and misinformation[M]. Latvia：NATO Strategic Communications Centre of Excellence，2021.

[33] PENNYCOOK G，RAND D G. Lazy，not biased：susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning[J]. Cognition，2019，188：39-50.

[34] SHAREVSKI F，ALSAADI R，JACHIM P，et al. Misinformation warnings：Twitter’s soft moderation effects on COVID-19 vaccine belief echoes[J]. Computers & Security，2022，114：102577.

[35] CLAYTON K，BLAIR S，BUSAM J A，et al. Real solutions

for fake news? measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media[J]. Political Behavior，2020，42（4）：1073-1095.

[36] BANERJEE S. Exaggeration in fake *vs.* authentic online reviews for luxury and budget hotels[J]. International Journal of Information Management，2022，62：102416.

[37] 邓胜利，顾一飞. 网络虚假健康信息研究综述：认知、行为与治理[J]. 图书馆杂志，2022，41（5）：14-22.

[38] KIM C，YANG S G. Like，comment，and share on Facebook：how each behavior differs from the other[J]. Public Relations Review，2017，43（2）：441-449.

[39] 位志广，宋小康，朱庆华，等. 基于随机森林的健康谣言分享意愿研究[J]. 现代情报，2020，40（5）：78-87.

[40] BASOL M，ROOZENBEEK J，BERRICHE M，et al. Towards psychological herd immunity：cross-cultural evidence for two prebunking interventions against COVID-19 misinformation[J]. Big Data & Society，2021，8：1-18.

[41] ROOZENBEEK J，VAN DER LINDEN S. How to combat health misinformation：a psychological approach[J]. American Journal of Health Promotion，2022，36（3）：569-575.

[42] LU C，HU B，LI Q，et al. Psychological inoculation for credibility assessment，sharing intention，and discernment of misinformation：systematic review and meta-analysis[J]. Journal of Medical Internet Research，2023，25：e49255.

作者简介

杨雪艳，女，硕士研究生，研究方向：信息管理和信息系统。
夏志杰，男，博士，教授，博士生导师，通信作者，研究方向：信息管理和信息系统，E-mail：xia_zhijie@163.com。

Influence of Interactive Pre-Intervention on False Information Perception and Behavior of Social Platform Users

YANG XueYan　XIA ZhiJie
(School of Management, Shanghai University of Engineering Science, Shanghai 201620, P. R. China)

Abstract: By combining logical vaccination-based pre-intervention measures with conversational artificial intelligence, the influence of intelligent system-driven interactive pre-intervention on social platform users' false information perception and behavior (like and forward) is verified, providing new ideas and methods for social platforms to combat false information. This study builds a framework model based on S-O-R theory, designs a control experiment, collects data by questionnaire survey, and explores the differential influence of intervention measures on users' false information perception and behavior through one-way ANOVA and mediation effect analysis. The results show that the intervention significantly reduces the perceived credibility, accuracy, and authenticity of false information and users' liking and forwarding behaviors. At the same time, perceived authenticity and perceived accuracy play a certain mediating role between intervention measures and user behavior. The study suggests that social platforms can reduce the negative influence of false information through pre-intervention measures, and take multiple factors into account when developing intervention strategies.

Keywords: Social Platform; False Information; Pre-Intervention; Individual Perception; Information Behavior; S-O-R Framework Model

（责任编辑：王玮）