

电商直播中对虚拟主播信任修复的机制与策略

谢昕儒

浙江理工大学理学院, 浙江 杭州

收稿日期: 2025年12月20日; 录用日期: 2025年12月30日; 发布日期: 2026年1月23日

摘要

虚拟主播在电商直播中的活跃度持续提升,但其引发的信任隐患亦逐步凸显。技术故障、推荐失误及“人设”崩塌等问题易触发独特信任危机,且相较于真人主播,消费者更倾向将虚拟主体失误归因为稳定内在缺陷,导致信任修复难度加大。本研究采用理论演绎与框架构建的研究方法,立足社会认知与人机交互双视角,旨在探索该信任修复的心理机制,解析信任违背类型及归因心理,构建以“归因干预”与“多维重建”为核心的修复模型。修复策略需覆盖能力、诚信、善意三维度,同时巧用虚拟形象“可塑性”,在提升交互公平感的同时规避过度拟真带来的伦理风险。

关键词

虚拟主播, 电商直播, 人机信任, 信任修复

Mechanisms and Strategies of Trust Repair toward Virtual Streamers in E-Commerce Live Streaming

Xinru Xie

School of Science, Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: December 20, 2025; accepted: December 30, 2025; published: January 23, 2026

Abstract

The engagement of virtual streamers in e-commerce live streaming continues to rise, yet the trust-related risks they pose are becoming increasingly evident. Issues such as technical failures, recommendation errors, and persona collapse can easily trigger unique trust crises. Compared with human streamers, consumers tend to attribute mistakes made by virtual entities to stable internal

flaws, making trust repair more challenging. This study adopts a methodology of theoretical deduction and framework construction, grounded in the dual perspectives of social cognition and human-computer interaction. It aims to explore the psychological mechanisms of trust repair in this context, analyze the types of trust violations and the psychology of attribution, and construct a repair model centered on “attributional intervention” and “multi-dimensional reconstruction”. Repair strategies should cover the three dimensions of competence, integrity, and benevolence, while skillfully leveraging the “plasticity” of virtual personas. This approach seeks to enhance the sense of interactive fairness while avoiding ethical risks associated with excessive anthropomorphism.

Keywords

Virtual Streamer, E-Commerce Live Streaming, Human-AI Trust, Trust Repair

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

随着 5G 网络、人工智能与计算机图形学技术的深度融合，电商直播已从新兴营销模式演进为主流消费场景[1]。在此背景下，虚拟主播作为一类由人工智能(AI)或“中之人”(背后真人演员)驱动的数字化身，凭借其可定制化的形象、24 小时不间断的播报能力以及降低真人主播“塌房”风险等优势，正快速渗透至直播带货领域[2][3]。从湖南卫视的“小漾”到各品牌自有的虚拟代言人，虚拟主播正重构着“人、货、场”中的“人”要素。

然而，技术的赋能亦伴随着新型的风险。虚拟主播的运作高度依赖算法、模型与数据，其交互行为本质上是预设规则与实时计算的产物。这使得其可能面临传统真人主播较少遭遇的信任危机：如因算法漏洞导致的产品信息播报错误(能力型违背)、因“中之人”不当言行引发的“人设崩塌”(诚信型违背)，或因交互脚本缺乏共情而被感知为纯粹的促销工具(善意型违背) [4] [5]。这些事件不仅直接损害销售转化，更可能动摇消费者对整个虚拟交互模式的信任基础。因此，一个核心问题亟待解答：当消费者对虚拟主播的信任因上述事件而受损后，应如何从心理学和人机信任的独特视角出发，进行有效且合乎伦理的信任修复？

本研究旨在系统回应这一问题。其理论意义在于：第一，将人际关系的信任修复理论拓展至人机交互，特别是虚拟主播这类具有高“社会临场感” [6] [7] 的拟人化主体情境，检验并发展经典理论的边界条件。第二，丰富人机信任理论体系，现有研究多关注信任的建立(如拟人化、专业性的影响) [8] [9]，而对信任破裂后的修复这一动态关键环节探讨不足 [10]。第三，为理解虚拟社会中人与非人类主体之间的复杂关系管理提供基于心理过程的深层解释。

其实践意义在于为电商平台、虚拟主播运营方及品牌方提供一套系统性的、基于心理机制的信任危机管理策略框架。这有助于将潜在的公关危机转化为展示责任感、提升系统透明度的契机，从而增强虚拟主播的商业可持续性，并最终保护消费者的合法权益与消费体验。

为实现上述目标，本文采用理论演绎与框架构建的研究方法。通过对心理学(归因、信任)、传播学(社会临场感、媒介等同)及信息系统(人机交互、技术接受)等多学科文献进行系统梳理与整合，并结合电商直播与虚拟主播的具体情境，构建一个针对性的理论分析模型，以揭示信任修复的内在心理机制并提出相应的策略体系。

2. 文献综述

2.1. 信任与信任修复的核心理论

2.1.1. 信任的多维度构成

Mayer 等人提出的经典人际信任模型认为，信任是对他人将做出对己方有利行动的积极预期，其形成基于对受信方三方面的感知：能力(技能与胜任力)、善意(动机上的利他关怀)和诚信(遵守道德原则)[11]。在人机信任情境中，该模型被广泛引用但需调整。对虚拟主播而言，能力体现为信息准确性、技术稳定性和交互流畅性[12]；善意虽源于程序设定，但通过共情化语言、个性化回应等拟人化设计来传递[13]；诚信则关联于其承诺的一致性、信息透明性及背后运营实体的道德操守[14]。消费者对虚拟主播的初始信任，正是基于对这些维度的初步评估而形成。

2.1.2. 信任修复的经典范式

信任修复的核心理论包括：(1) 归因理论：受害者对违背原因的推断(内在/外在、稳定/不稳定、可控/不可控)深刻影响其愤怒程度与修复难度。内在、稳定、可控的归因会导致更严重的信任损伤[15]。(2) 违反期望模型：违背事件破坏了受害者对受信方行为模式的期望，修复需重构一致性预期。(3) 修复策略双路径：通常分为言语修复(如道歉、解释)与实质修复(如赔偿、制度整改)。研究指出，对于能力型违背，实质修复更有效；对于诚信型违背，真诚的言语修复结合实质补偿至关重要[16]。

2.2. 人机信任的特殊性

人机信任的建立与修复有其独特性。首先，用户更强调系统的可靠性与可预测性，任何非预期的故障或错误都易被放大。其次，透明度成为关键，算法“黑箱”会加剧不信任感[12]。再次，用户对机器的可控感(如是否有中止或反馈渠道)显著影响信任水平。此外，“恐怖谷效应”提示，当虚拟形象高度拟人却仍有非人瑕疵时，可能引发本能的不适与疏离，使信任基础更为脆弱[8]。

2.3. 虚拟主播的特性与信任建立

2.3.1. 虚拟主播的界定与分类

虚拟主播可分为 AI 驱动型(完全由算法生成内容与交互)与“中之人”驱动型(真人演员通过动作捕捉等技术实时操控虚拟形象)。它们共同具有“拟人化”与“媒介化”双重属性：既通过外观、声音、行为模仿人类以唤起社会反应，又始终作为技术媒介被用户所认知[17] [18]。

2.3.2. 影响信任建立的关键心理因素

虚拟主播之所以能打动观众，一个关键心理机制在于“社会临场感”——即观众透过屏幕，依然能感受到对方仿佛真实存在的程度[19]。这种“在场”的错觉，是建立情感联结与初步信任的基石[20]。然而，隔着数字媒介的交流终归滤掉了大量现实社交的线索，于是观众只得抓住有限的信号来做出判断：主播是否专业？回应是否及时？形象与言行是否一致？这些碎片便成了评估其可信度的主要依据。这引出了一个问题：拟人化。赋予虚拟形象以人的特征，犹如一把双刃剑。适度为之，能拉近距离，唤起信任；可一旦过度拟真，却未能匹配相应的智能或共情能力，观众的期待便会落空，反而可能导致更深的疏离与信任崩塌。

3. 电商虚拟主播的信任违背类型与消费者心理归因

3.1. 信任违背的主要类型

结合电商直播场景与虚拟主播特性，其信任违背主要分为以下三类。

3.1.1. 能力型违背

指虚拟主播未能胜任其功能角色。包括：1) 知识性错误：播报的产品参数、功效等信息不实；2) 技术性故障：出现模型穿模、音画不同步、宕机等“穿帮”现象；3) 交互性缺陷：对用户提问题答非所问、反应迟钝，暴露其AI能力的局限。

3.1.2. 诚信型违背

指虚拟主播或其运营方被感知为有意欺骗。包括：1) 虚假宣传：夸大产品效果，隐瞒缺陷；2) 人设崩塌：“中之人”的负面言行被曝光，导致其驱动的虚拟角色形象受损；3) 算法偏见：推荐逻辑被质疑存在不公正，如“大数据杀熟”。

3.1.3. 善意型违背

指虚拟主播被感知为缺乏对消费者的关怀。包括：1) 共情缺失：面对消费者投诉或负面反馈时，回应模板化、冷漠；2) 利益导向过于明显：所有互动均被导向促销，缺乏真诚交流。

3.2. 消费者对虚拟主播失信行为的心理归因过程

理解消费者在信任违背发生后的心理归因过程，是构建有效修复策略的认知起点。一旦失信事件发生，用户会自发启动一个多阶段的归因分析。首先是对归因对象的辨识：责任应归属于作为前端的虚拟形象本身、其运营团队、技术提供商，还是委托品牌方？虚拟主播往往成为所有后端复杂性的单一投射界面，极易承担“替代性责任”。

随后是对归因维度的评估，这一过程沿着三个关键轴向展开：其一，判断原因是源于主体内在的特性缺陷(如算法设计瑕疵)，还是外部情境因素(如临时性技术故障)；其二，评估该原因是稳定的、反映固有缺陷的，还是不稳定的、偶然出现的；其三，考量该结果在事前是否可被责任方预见并控制。

尤为关键的是，虚拟主体的本质属性在此过程中扮演着核心的调节角色。由于其内在决策逻辑与情感模拟机制具有高度的“不透明性”[10]，在信息缺失的情境下，用户更容易形成内在化、稳定化的负面归因。例如，一次产品信息播报错误，可能不会被解释为一次性的数据源差错，而是被归结为“该人工智能系统本质上缺乏可靠性”，甚至被质疑为预设的商业欺骗策略。这种归因偏差会急剧放大信任损伤的强度与持久性。因此，信任修复的初始干预点，必须精准地指向这一认知环节，致力于对自发形成的消极归因模式进行重构与引导。

4. 虚拟主播信任修复的心理机制与策略构建

4.1. 信任修复的机制模型

基于以上分析，本文提出一个针对虚拟主播的信任修复的机制模型(图1)。其核心逻辑是：有效的修复策略通过作用于消费者的认知重评(改变归因)与情感安抚(缓解负面情绪)，进而重建对虚拟主播在能力、诚信、善意三个维度上的积极信念(输出)。在整个过程中，“感知真实性”(即消费者认为修复行为是真诚、非操纵的)和“交互公平性”(即消费者认为其关切被严肃对待、过程公正)扮演着关键的中介角色。若修复行为被视为公关表演或推卸责任，则所有策略将失效。

4.2. 基于归因管理的修复策略

核心目标是引导消费者从内在、稳定、可控的负面归因，转向外在、不稳定或可控性较低的归因。

4.2.1. 针对外在归因的引导

对于由外部因素引发的故障，必须及时、具体、透明地进行解释。例如，直播中出现卡顿，虚拟主播

或屏幕提示应明确告知：“由于当前区域网络波动，我的响应出现了延迟，技术团队正在紧急修复。”这能将问题与虚拟主播的核心能力暂时剥离。

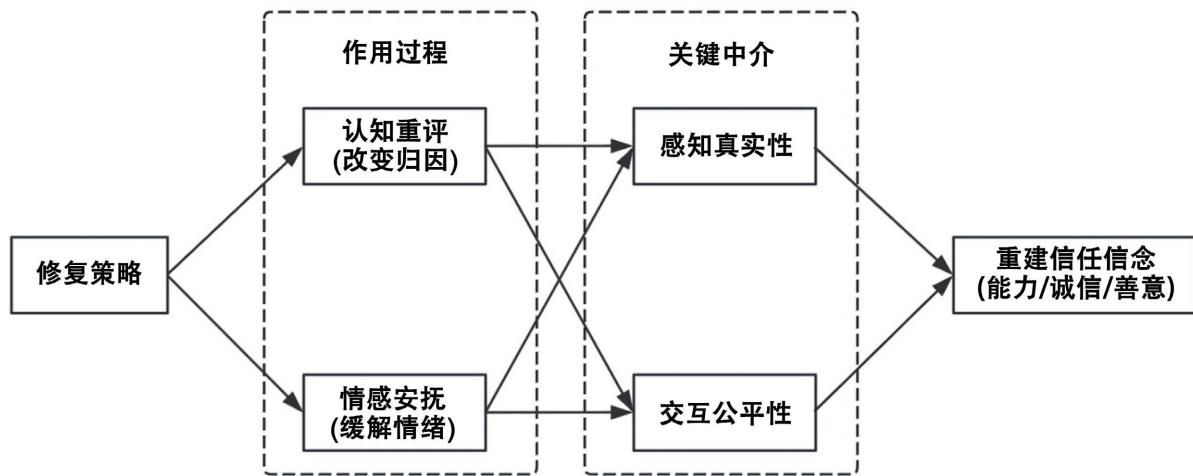


Figure 1. Mechanism model of trust repair
图 1. 信任修复的机制模型

4.2.2. 降低内在、稳定归因

当问题确实源于自身时，策略重点在于将“缺陷”重构为“可进化过程”。例如，在发生推荐错误后，虚拟主播可以在后续直播中提及：“感谢大家上次的指正，我们的知识库已经更新了该产品的信息。AI 和人类一样，也在不断学习和成长。”这利用了 AI 的“可学习”特性，将一次失误转化为展示进步诚意的机会。

4.3. 基于信任维度重建的具体策略

4.3.1. 能力维度修复

1) 即时纠错与透明化：一旦发现口误或信息错误，立即以显著方式纠正，并简要说明错误来源(如“刚才引用了旧版资料，正确信息是……”)。2) 引入权威第三方背书：在涉及专业产品时，安排与真人专家、品牌工程师的连麦环节，让虚拟主播扮演“好学助手”或“高效展示者”角色，借助真人权威弥补自身可信度[21]。3) 展示持续优化证据：定期通过“开发者日志”或直播小环节，以通俗方式介绍 AI 模型的迭代升级，如“新版模型提升了理解复杂问题的能力”。

4.3.2. 诚信维度修复

1) 符合人设的真诚道歉：道歉需符合虚拟角色的设定风格，但核心信息必须清晰、负责。避免过于拟真引发“情感欺骗”感[17]。例如，一个活泼的虚拟偶像可以说：“让小主们失望了，我这次没做好功课，非常抱歉！”同时，道歉主体应明确为运营方。2) 实质性补偿：提供无门槛优惠券、退款、赠品等，这是修复诚信感知最有力的行为语言[16]。3) 承诺并公示流程改进：宣布并后续公示已采取的改进措施，如“已引入三重人工审核机制确保产品信息准确”。

4.3.3. 善意维度修复

1) 增强交互的共情性与响应性：在事件后的直播中，主动关怀受影响的用户，如点名回应其后续问题，使用更多共情语言。2) 增加社会性承诺：将部分因事件带来的流量或收益用于公益捐赠，并公示过程和结果，塑造超越商业利益的“社会公民”形象，提升善意感知。

4.4. 虚拟主播特有的修复策略与伦理考量

虚拟主播的“数字本体”特性，如同一把双刃剑，既开辟了独特的修复路径，也暗藏着伦理的暗礁。其可塑性为修复提供了创造性空间。一方面，运营者可主动构建“修复叙事”——通过一则以虚拟主播为主角的短视频或特别直播，以第一人称视角诚恳回溯事件，将技术说明与改进承诺融入故事脉络，从而将危机应对本身转化为一种富有张力的内容体验。另一方面，可适时揭开部分“后台”帷幕，例如以“技术探秘”或“制作日志”等形式，有限度地展示形象生成、驱动与互动的幕后流程。这种有控制的透明化，既能消解因未知而产生的疑虑，也有助于将信任从单纯的“人格化”投射，部分转向对其背后技术逻辑的理性认同。

然而，这种可塑性也亟需伦理的缰绳。其中最需警惕的是“拟人化道歉”的滥用。当虚拟形象以高度拟真的姿态、充满情感张力的方式进行道歉，若缺乏后续实质性的纠正与补偿，极易被感知为一种更精巧的操控，即利用人们对“人性化”交互的情感期待进行欺骗。因此，修复策略的伦理底线必须明确：其一，责任主体必须清晰，虚拟角色不能成为运营方规避现实责任的“替身”；其二，须始终保持技术透明，避免刻意营造 AI 拥有超越其本质的情感与主观意识的不实幻想，防止信任建立在错误的认知基础上。

5. 讨论

5.1. 研究发现总结

本文系统论证了电商直播中虚拟主播信任修复问题的特殊性与复杂性。研究指出，虚拟主播的信任危机不仅源于事件本身，更源于消费者因其“虚拟性”而更容易产生的内在、稳定的负面归因。因此，有效的修复是一个系统的心理工程，其核心机制在于通过策略性沟通与行动，干预消费者的归因过程，并在能力、诚信、善意三个维度上同步进行信念重建。成功的修复必须充分利用虚拟主播的媒介特性(如可塑性、可展示性)来提升透明度和公平感，同时严格遵守不误导、不逃避的伦理准则。

5.2. 理论贡献与实践启示

5.2.1. 理论贡献

首先，本研究将归因理论置于人机信任修复的中心位置，明确了其作为修复过程“认知开关”的关键作用，弥补了以往研究多从策略效果出发而较少深入心理过程的不足。其次，提出的整合模型为分析虚拟主体信任修复提供了一个涵盖前因(归因)、中介(感知真实性、公平性)和结果(信任维度)的框架，促进了理论的系统化。最后，研究特别强调了在“感知可控性”普遍较低的人机交互中，主动提供“透明度”是重建控制感和信任的特殊价值路径。

5.2.2. 实践启示

为运营方提出了全周期的信任管理建议：事前，通过用户教育合理设定对虚拟主播能力的期望，避免过度营销；事中，建立包括归因快速判断、多策略组合响应、统一信息出口在内的危机响应流程；事后，将修复承诺长期化、可视化，建立常态化的沟通机制(如定期透明度报告)。长期而言，构建开放的“算法/运营透明度”体系，是预防信任危机、筑牢信任基石的终极策略。

5.3. 研究局限与未来展望

本文作为一项理论构建研究，主要局限在于所提模型与策略尚未经过大样本实证数据的严格检验。未来研究可以从以下方向深入：第一，开展实验研究或问卷调查，量化比较不同类型修复策略对不同信

任违背类型的有效性。第二，进行跨文化比较研究，探讨文化维度(如个人主义/集体主义)如何影响消费者对虚拟主播失信事件的归因及对修复策略的接受度。第三，聚焦于某一特定高发场景(如“中之人”争议)进行深入的案例研究与过程追踪，提炼更具情境性的修复范式。

6. 结论

在电商直播构建的数字消费新景观中，虚拟主播作为联结品牌与消费者的前沿界面，其信任关系的稳固与否直接关乎商业价值的实现。本研究揭示，对虚拟主播的信任修复，远不止于一次技术故障的排除或一则公关声明的发布，其深层是对消费者在遭遇人机交互失效后所产生的复杂心理过程——尤其是归因过程——的理解、尊重与巧妙引导。成功的修复策略，需要有机融合对社会认知规律的把握、对人机交互特点的运用以及对商业伦理底线的坚守。唯有如此，方能在虚拟世界中，将一次不可避免的信任危机，转化为一次深化关系、彰显责任、并最终提升整个系统可信度与鲁棒性的宝贵契机。

参考文献

- [1] 王宝义. 直播电商的本质、逻辑与趋势展望[J]. 中国流通经济, 2021, 35(4): 48-57.
- [2] 申凌. AI 虚拟主播在新闻报道中的应用与发展[J]. 媒体融合新观察, 2023(5): 24-27.
- [3] 何欣宇, 袁星洁. 智媒时代 AI 虚拟主播在电视媒体中的应用——以湖南卫视数字主持人“小漾”为例[J]. 新闻前哨, 2023(12): 7-9.
- [4] 宋婷. 虚拟新闻主播实践中的技术困境与伦理问题研究[D]: [硕士学位论文]. 重庆: 重庆大学, 2023.
- [5] 李颖. 虚假数据坑苦商家信任修复需多方合力[J]. 中国质量万里行, 2025(9): 6-7.
- [6] 肖静. 基于社会临场感视角的电商直播平台消费者购买意愿研究[D]: [硕士学位论文]. 桂林: 桂林理工大学, 2020.
- [7] 李明伟, 魏亚楠. 电商直播平台用户持续使用和推荐意愿研究——基于直播社会临场感视角的分析[J]. 价格理论与实践, 2024(3): 174-178.
- [8] 覃周瑶. AI 虚拟主播形象拟人化程度对消费者购买意愿的影响研究[D]: [硕士学位论文]. 南宁: 广西大学, 2025.
- [9] 李茸, 马宝龙, 张鹏, 等. 电商 AI 主播特征影响消费者购买意愿的机制——基于技术感知与社会互动双重视角[J]. 北京理工大学学报(社会科学版), 2025, 27(1): 127-144, 187.
- [10] 计康凌. 虚拟主播拟人化程度对信任修复的影响研究[D]: [硕士学位论文]. 武汉: 中南财经政法大学, 2023.
- [11] Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995) An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20, 709-734. <https://doi.org/10.2307/258792>
- [12] 刘漫森. AI 主播信任度影响因素研究[D]: [硕士学位论文]. 保定: 河北大学, 2023.
- [13] 谭瑶. 智媒时代下“AI 合成主播”情感化设计研究[D]: [硕士学位论文]. 吉林: 东北电力大学, 2024.
- [14] 金毓. 电商直播中消费者信任感知对购买意愿的影响研究[J]. 中国商论, 2022(19): 40-43.
- [15] 王金妹. 主播带货负面事件对消费者购买意愿的影响及其修复研究[D]: [硕士学位论文]. 烟台: 烟台大学, 2023.
- [16] 谢雪梅, 刘璐瑶. “直播 + 电商”模式下多元互动信任修复机制研究[J]. 北京邮电大学学报(社会科学版), 2023, 25(2): 56-67.
- [17] 赵瑜, 李孟倩. 拟人化趋势下的虚拟主播实践与人机情感交互[J]. 现代传播(中国传媒大学学报), 2023, 45(1): 110-116.
- [18] 钟丁静, 吴风, 邱锐. 拟人化与智能化: AI 主播媒介性与人机信任关系建构的实证研究[J]. 国际新闻界, 2025, 47(2): 49-71.
- [19] 徐新, 杨燕. 直播带货社会临场感对消费者购买意愿的影响——基于顾客粘性中介与主播类型调节作用[J]. 商业经济研究, 2024(4): 83-86.
- [20] 雍巧云. 电商直播情境下社会临场感、消费者信任和持续购买意愿的关系研究[D]: [硕士学位论文]. 北京: 中央民族大学, 2021.
- [21] 王爽. 电商直播中虚拟主播对消费者购买意愿影响研究[D]: [硕士学位论文]. 福州: 福建师范大学, 2023.