E-Commerce Letters 电子商务评论, 2026, 15(2), 38-44
Published Online February 2026 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2026.152126

Hans汉斯

AI生成内容时代的在线评论生态危机与协同治理机制研究

樊佳琪，朱　迅

江苏大学管理学院，江苏　镇江

收稿日期：2026年1月5日；录用日期：2026年1月19日；发布日期：2026年2月4日

摘　要

生成式人工智能(AIGC)技术的爆发式发展，正以前所未有的深度和广度重塑数字内容的生产与传播范式。本研究旨在系统探究AIGC诱发的在线评论生态危机及其协同治理机制，通过系统剖析人工智能生成内容对电商评论生态在信息质量、消费者信任、市场效率及平台治理四个维度构成的系统性冲击，揭示其通过“信息污染”引发“信任危机”、最终导致“市场失灵”的内在机理。基于此，本文提出了一个涵盖技术规制、平台责任、法律监管和社会监督的协同治理框架，并设计了具体可行的治理机制。本研究认为，应对AIGC时代的评论生态危机，必须超越单一主体、单一维度的传统治理思维，构建多元主体协同、多维度联动的新型治理范式。

关键词

生成式人工智能，在线评论，信息生态危机，协同治理

Research on the Online Review Ecological Crisis and Collaborative Governance Mechanisms in the Era of AI-Generated Content

Jiaqi Fan, Xun Zhu

School of Management, Jiangsu University, Zhenjiang Jiangsu

Received: January 5, 2026; accepted: January 19, 2026; published: February 4, 2026

文章引用: 樊佳琪, 朱迅. AI 生成内容时代的在线评论生态危机与协同治理机制研究[J]. 电子商务评论, 2026, 15(2): 38-44. DOI: 10.12677/ecl.2026.152126

Abstract

The explosive development of generative artificial intelligence (AIGC) technology is reshaping the production and dissemination paradigms of digital content with an unprecedented depth and breadth. This study aims to systematically explore the online review ecological crisis induced by AIGC and its collaborative governance mechanisms. By comprehensively analyzing the systematic impacts of AI-generated content on the e-commerce review ecosystem across four dimensions—information quality, consumer trust, market efficiency, and platform governance—it reveals the inherent mechanism through which "information pollution" triggers a "trust crisis" and ultimately leads to "market failure." Based on this, this paper proposes a collaborative governance framework encompassing technological regulation, platform responsibility, legal supervision, and social oversight, along with designing specific and feasible governance mechanisms. This study argues that to address the review ecological crisis in the AIGC era, it is essential to transcend traditional governance thinking centered on a single entity or dimension and instead construct a new governance paradigm characterized by multi-entity collaboration and multi-dimensional interaction.

Keywords

Generative Artificial Intelligence, Online Reviews, Information Ecological Crisis, Trust Mechanism, Collaborative Governance

1. 引言

在线评论作为数字经济时代的重要社会技术系统，承载着多重社会功能：在微观层面，它是消费者购买决策的关键依据；在中观层面，它构成了商家声誉数字化的核心载体；在宏观层面，它影响着市场资源配置的效率与公平。主流电商平台已建立起基于评论的复杂生态系统，包括评分系统、排名算法、用户互动机制等，在降低信息不对称、辅助购买决策、构建买卖信任等方面发挥着不可替代的作用，其价值根基在于“真实用户体验”的汇聚与共享[1]。自 2022 年以来，以 ChatGPT 为代表的生成式人工智能技术实现了突破性进展，标志着人工智能从“感知理解”向“创造生成”的范式转变。然而，生成式人工智能技术的突破性发展，特别是大语言模型的通用化与低成本化，使得大规模、高质量、低成本的 AIGC 生产成为可能[2]。这一技术浪潮正剧烈冲刷着电商评论生态的基石：AI 可以轻易模仿人类口吻，批量生成以假乱真的“体验式”评论，其规模与效率远超传统“水军”[3]。因此，本研究在梳理 AI 生成内容与在线评论的研究基础上，运用信息生态理论、信任与信号理论等理论，提出的协同治理框架，为理解多元主体在复杂技术环境下的合作治理机制提供了理论模型。

2. 文献综述与分析框架构建

2.1. 传统在线评论研究

已有研究主要围绕三个主题展开：一是评论的影响力研究，聚焦于评论数量、评分、情感倾向等对产品销量[4]、消费者决策的量化影响[5]；二是评论的有用性与可信度研究[6]，探讨文本特征(如长度、细节)、发布者特征及社会互动(如“有用”投票)如何影响感知可信度；三是虚假评论的检测与治理研究[7]，

多从行为模式、语言风格、网络关系等角度利用机器学习进行识别。这些研究奠定了坚实基础，但其前提是假设评论主体为“人”，治理对象是“低质量人工内容”[8]。

2.2. AIGC 的发展现状

AIGC 指利用人工智能算法，通过学习大规模数据集中的模式和规律，自动生成符合特定要求的内容。与传统的规则驱动或检索式内容生成不同，AIGC 基于深度学习模型，具有创造性、适应性和规模化的特点。在本研究中，AIGC 特指应用于电子商务场景的内容生成，包括但不限于商品评价、使用体验描述、问答回复等。例如，电商领域商家可以利用 AIGC 快速生成商品宣传图与模特换装图等，以此大幅降低摄影与模特成本[9]。AIGC 的核心特征是算法主导的内容生产过程，其内容质量取决于训练数据的广度、模型架构的先进性和提示工程的精细程度。现有研究多呈点状分布，或聚焦技术单点突破，或讨论宏观伦理原则，缺乏将 AIGC 置于特定商业生态(如在线评论)中[10]，进行“冲击－响应”的系统性分析。

2.3. 分析框架的构建

为弥补上述缺口，本文构建一个二维分析框架，纵向维度为“冲击层面”，涵盖从表层的信息质量，到中观的消费者认知与信任，再到宏观的市场运行效率，最后触及支撑系统的平台治理规则。这四个层面由表及里，逐级传导。横向维度为“治理主体”，涵盖技术提供方与使用方、平台企业、政府监管机构、行业组织与社会公众。本框架旨在阐明，AIGC 的冲击是系统性的，必须依靠多元主体的协同响应，在不同冲击层面施加治理干预，方能实现生态的再平衡。

3. 生成式 AI 对评论生态的多维重塑

3.1. 信息质量维度

尽管传统虚假评论依然存在，然而受限于人力与成本因素，其影响范围相对较为局限。相关研究表明，传统虚假评论在整体评论体系中所占比例通常维持在较低水平。并且，部分虚假评论因带有显著的非理性情感宣泄特征，例如过度地褒扬或谩骂，相对而言较易被识别[11]。随着 AIGC (人工智能生成内容)的介入，信息“污染”状况发生了本质性的转变。其一，呈现出规模化与同质化特征。以某知名美妆产品评论区为例，在引入 AIGC 后，短时间内出现了数万条语义通顺、情感积极的评论，这些评论在表述上高度雷同[12]，如“这款粉底液遮瑕效果超棒，持妆一整天都没问题”，形成了典型的“信息堰塞湖”现象。真实用户原本多元的声音，如对产品质地、色号适用性等方面的不同反馈，被淹没其中。其二，存在属性空心化与细节幻觉问题。AIGC 虽具备编造看似合理细节的能力，例如“冬天续航缩水”这类表述，但这些细节并非源于真实体验，而是基于训练数据中概率关联的拼接组合。这种特性可能导致商品属性描述与实际情况严重偏离，进而形成误导性更强的“细节幻觉”。

3.2. 消费者信任维度

评论系统的有效性构建在“默认大多数评论者秉持诚实态度”这一脆弱的社会契约基础之上。AIGC 的出现动摇了这一基石。其一，AIGC 引发了消费者对评论系统的普遍性质疑。当消费者意识到评论可能由人工智能生成时，其对整个评论板块的初始信任程度会普遍降低，进而陷入“先疑后证”的心理认知模式。以酒店点评场景为例，一旦消费者感知到评论是由 ChatGPT 生成，便会认为此类评论的有用性更低、可信度更差、真实性不足。并且，这种负面效应不会因评论是好评或差评，亦或是标注有“资深用户”身份提示而有所改变[13]。其二，AIGC 导致评论验证机制失效以及消费者决策瘫痪。在传统情境下，消费者会通过交叉比对不同评论内容、探寻评论中的瑕疵或极端评价等方式，对评论进行综合分析判断。

然而，高度拟真且逻辑自洽的 AIGC 评论使得这种基于文本内部一致性以及社会线索的验证方法难以奏效。一旦消费者怀疑存在“人为操纵评论或人工智能生成评论”的情况，其对整个平台的信任会显著崩塌。这种信任的崩塌会进一步削弱评论对消费者决策的参考价值，甚至可能导致消费者放弃对评论系统的依赖[14]。

3.3. 市场效率维度

健康的市场依赖于准确的质量信号来实现资源的有效配置。AIGC 的泛滥严重扭曲了这一关键信号。第一，逆向选择风险加剧。诚信卖家依靠真实口碑积累信誉，但在成本方面处于劣势；而不法卖家则可利用低成本 AIGC 迅速营造“虚假繁荣”的局面，获取不公平的流量与销量优势。一旦竞争对手大量造假，其他商家会被强烈诱导跟进刷评，否则难以维持竞争力，形成“囚徒困境”式集体堕落[15]。第二，价格与质量的关联性断裂。当评论无法真实反映产品质量时，价格作为另一核心市场信号也失去了参照依据，消费者难以做出效用最大化的选择，市场资源配置效率显著降低。

3.4. 平台治理维度

电商平台多年来依赖的反作弊系统(例如识别异常发布时间、购买关系、文本模式等)在面对高质量 AIGC 时迅速“钝化”。AI 生成的内容在语言规范性、情感分布等方面可能比部分真实评论更符合优质评论的特征，在生成产品评论时，更具分析性、情感更饱满、形容词更丰富，但整体结构更规整，在自动评价中被判定为“高质量文本”的比例很高[16]，这使得基于规则和传统机器学习的过滤器难以发挥作用。同时，平台陷入“检测–生成”的军备竞赛，治理成本急剧攀升。此外，平台商业目标(追求商品交易总额增长与内容丰富度)与生态治理目标(保障内容真实性)之间的内在矛盾，在 AIGC 时代被进一步放大。

4. 协同治理进程中所面临的阻碍因素

4.1. 技术层面的“矛与盾”动态博弈

AIGC 检测技术与生成技术本质上构成了一场持续的动态较量。目前，检测模型的发展进程往往滞后于最新生成模型的更新迭代速度，这种技术发展的不同步性引发了两类关键风险，即“误杀”现象(将真实用户评论错误地判定为 AI 生成内容)与“漏杀”现象(未能有效识别出 AI 生成内容)。现有的文本检测工具在准确性方面存在显著局限，误报率和漏报率均较为突出，整体呈现出“既缺乏精准度又缺乏可靠性”的特征。并且，这些检测工具存在明显的判定偏向，更倾向于将文本判定为“人类撰写”，这一倾向导致大量 AI 生成的文本被漏检。同时，对于经过改写或翻译处理的 AI 文本，现有检测工具几乎完全失效[17]。实现高精度的检测往往需要庞大的算力以及海量的数据作为支撑，这不仅导致成本居高不下，而且难以达成实时、全量的覆盖效果。这种技术层面所存在的不确定性以及高昂成本，已然成为部署有效治理工具的首要阻碍因素。

4.2. 平台层面的利益冲突与责任模糊

平台企业作为生态守门人，面临多重压力。一方面，严厉治理 AIGC 可能短期内影响平台内容的“丰富度”与部分商家的活跃度；另一方面，若放任不管，长期将损害生态公信力，危及根本。此外，平台在 AIGC 管理中的法律地位与责任边界尚不清晰——是承担“严格审查义务”还是“事后合理注意义务”？这一定位的模糊影响了平台投入治理资源的决心与尺度。

4.3. 法律规制层面的滞后性与全球协调难题

现有的法律体系主要是围绕“虚假宣传”“商业欺诈”等概念构建而成的，然而，当把“未声明且具

有误导性的 AIGC”纳入这一范畴时，在取证以及定性方面存在诸多困难。AIGC 的“虚假”主要体现在内容来源的虚假性，而非陈述事实的完全虚构，这给法律的适用带来了严峻挑战。与此同时，由于数字服务具有全球性的特征，任何一个单一司法管辖区所制定的规制措施，都可能面临效力外溢或者管辖权冲突等问题，因此，迫切需要进行国际间的协调与合作。

4.4. 社会认知层面的素养鸿沟

广大消费者与中小商家普遍缺乏辨识 AIGC 的意识和能力，处于信息劣势地位。据调查，在电商与点评场景中，参与者区分人写评论与 AIGC 评论的正确率仅略高于随机猜测，AI 评论经常被误认为真人评论[18]。用户主观上既难以判断来源，又经常高估自己的识别能力。这种数字素养差距使得社会监督力量难以充分发挥作用，消费者无法有效监督评论真实性，中小商家也难以在自身经营中避免受到 AIGC 的误导。同时，加剧了治理对技术和平台的单向依赖程度，无法形成有效的治理合力，影响了整体治理效果

5. 面向 AI 生成内容的在线评论协同治理机制构建

5.1. 技术识别与源头规制机制

在技术识别方面，着力发展深度检测技术。鼓励科研机构与企业投入资源，研发基于大语言模型内在特征的新型检测算法，以此提升算法对抗性攻击的鲁棒性，增强检测的准确性与稳定性。推行强制水印与元数据标识制度，推动在相关技术标准中嵌入隐形数字水印或可验证的元数据，实现 AIGC 的源头追溯与一键识别。同时，开发用户端辅助工具，平台或第三方可提供轻量化的浏览器插件或移动端应用程序，对可疑评论进行实时风险提示，辅助用户辨别信息真伪。

技术治理应从侧重事后检测向强化源头控制延伸，构建多层防御体系：第一层为源头标识与数字水印层。推动建立 AIGC 内容的强制标识制度，要求所有可用于商业内容生成的 AI 系统，在输出内容时嵌入不可移除的元数据水印或显性标识(如“[AI 生成]”标签)。鉴于数字服务的跨国性，这需要在国际层面协调技术标准，确保标识的统一性和互操作性，以便在全球范围内有效识别 AIGC。同时，鼓励科研人员开发更先进的数字水印技术，使水印能够抵御常见的内容修改操作，如裁剪、压缩、格式转换等，保障标识的持久性和可靠性。

第二层是多模态深度检测技术层。超越单纯的文本分析范畴，构建融合多维度数据的检测系统。综合考量文本特征(包括语法模式、情感分布、信息密度等)、行为特征(涵盖发布频率、账户历史、互动模式等)、交易特征(如购买验证、物流数据等)、网络特征(例如发布 IP、社交关系等)等多重信号。运用图神经网络等先进技术，深度挖掘数据之间的关联关系，识别隐藏在复杂关系网络中的操纵行为，提高检测的全面性和精准性。

第三层是基于区块链的存证与验证层。探索建立去中心化的真实体验存证系统，鼓励消费者将关键购物凭证(如订单哈希、支付证明、开箱视频指纹等)与评论关联上链。区块链技术具有不可篡改的特性，为真实评论提供了可验证的技术背书，同时通过加密等技术手段保护用户隐私。平台可以对经过验证的评论给予更高权重，在排序算法中优先展示，引导消费者关注真实可靠的评价信息。

5.2. 平台主体责任与算法透明机制

设计“真实性加权”排序算法，改进现有的评论排序机制。为通过强验证(如视频评价、已购用户长图文等)的内容赋予更高权重，使其在搜索结果和展示页面中优先展示，提高真实评论的曝光度。

建立 AIGC 分级披露制度，要求商家或内容提供者明确标注 AI 辅助或生成的内容。对于未标注的行

为设定明确的处罚规则，如降权、流量限制等，促使商家和内容提供者遵守规定，保障消费者的知情权。

创新真实性激励体系，通过积分、荣誉、流量激励等方式，鼓励用户发布高质量真实评价。从供给端增加真实评价的数量，稀释 AIGC 在评论中的比例，营造健康、真实的评论环境。

5.3. 法律规制与监管执法机制

完善法律定义与责任体系，建议修订《电子商务法》《反不正当竞争法》《广告法》等相关法律，明确将“利用人工智能技术生成虚假商品声誉信息”列为违法行为，为打击此类行为提供明确的法律依据。在责任认定上，建立多层次责任体系。内容生成者作为直接参与者，承担直接责任；平台在知情不处理或未采取合理措施时，需承担相应的监管责任；工具提供者若故意设计用于违法用途或明知违法用途仍提供技术支持，应承担连带责任，形成全方位的责任约束机制。

创新监管工具与方法，监管机构应采用与技术发展相适应的监管手段。推行监管沙盒制度，在可控环境中测试新的治理方案，评估其有效性和可行性，降低监管风险。采用基于风险的差异化监管策略，对大型平台、高风险领域实施重点监管，提高监管效率和针对性。加强技术监管能力建设，建立专业的技术团队，与学术界、产业界保持密切合作，及时掌握技术发展动态，提升监管的技术水平和应对能力。

强化跨境执法协作，鉴于 AIGC 评论的跨国性特征，加强国际执法合作至关重要。通过双边和多边协议，建立信息共享、调查协助、联合行动等机制，加强各国在打击 AIGC 虚假评论方面的协作与沟通。在国际标准组织中积极参与相关标准制定，推动全球治理规则的协调统一，共同应对 AIGC 带来的挑战。

5.4. 素养提升与社会监督机制

在数字素养与风险教育方面，将 AIGC 风险识别纳入公民数字素养教育体系[19]。通过公益广告、平台提示、学校教育等多种渠道，广泛普及 AIGC 评论的特征和识别技巧。教育内容应注重实操性，例如教导消费者如何识别模式化表达、如何寻找验证信息、如何使用辅助检测工具等，提高消费者的数字素养和风险识别能力，使其能够在复杂的网络环境中准确辨别信息真伪。

优化用户举报与反馈系统，平台应建立便捷、透明、有效的举报渠道。优化举报流程，减少用户操作成本，提高用户举报的积极性。提高处理透明度，及时向用户反馈处理结果，让用户了解举报的进展和成效。建立举报奖励机制，对提供有效线索的用户给予适当激励，如积分、优惠券等，鼓励更多用户参与监督。同时，鼓励发展专注于消费者权益保护的社会组织和媒体，发挥舆论监督作用，形成全社会共同参与的监督格局。

发展替代性信任机制，在评论系统之外，培育多元化的信任建立机制。鼓励发展基于真实体验的深度评测社区，为消费者提供专业、深入的产品评测信息；支持专业媒体和 KOL 开展客观的产品评测，借助其专业知识和影响力，引导消费者做出理性决策；探索基于社交关系的推荐网络，利用消费者之间的信任关系传递真实的产品信息。这些替代机制不仅为消费者提供了更多选择，也构成了对评论系统的竞争性约束，促使评论系统不断改进和完善。

6. 结论与展望

本文系统论证了生成式 AI 对电子商务评论生态的冲击绝非局部或技术性的，而是从信息源污染出发，传导至消费者信任崩塌，最终导致市场机制扭曲与平台治理失灵的系统性风险。其本质是一场由技术引发的、关乎数字经济基础信任体系的危机。应对此危机，任何单点突破均不足以治本，必须采用系统思维，实施技术侦测、平台规则、法律规制与社会共治紧密配合的协同治理。对于平台企业而言，应超越短期流量思维，将生态真实性作为长期核心资产进行投资，主动创新治理工具与规则。对于监管部门，需秉持敏捷与包容审慎原则，加快立法进程，明确底线规则。对于消费者与社会，应积极提升自身

数字能力，参与生态监督。

参考文献

[1] 杜茜. 在线评论真实性对消费者购买意愿的影响研究[J]. 商业经济研究, 2025(24): 75-78.

[2] 关乐宁, 徐凌验. 通用目的技术视角下新一代人工智能的作用机理与治理体系[J]. 系统工程理论与实践, 2024, 44(1): 245-259.

[3] 张文, 王强, 马振中, 等. 在线商品虚假评论发布动机及形成机理研究[J]. 中国管理科学, 2022, 30(7): 176-188.

[4] 汪宇萌, 刘莉. 非标准住宿形态下顾客感知价值分析——以文化主题酒店在线评论为例[J]. 商业观察, 2025, 11(33): 86-90.

[5] 黄欢. 农产品短视频评论对消费者购买决策的影响研究——基于扎根理论的探索[J]. 商展经济, 2025(21): 120-127.

[6] 崔登峰, 李锦秀, 王海忠. 什么样的评论更可信? 心理模拟视角下在线评论类型对感知在线评论可信性的影响研究[J]. 商业经济与管理, 2022(2): 29-42.

[7] 张国防, 袁国强, 赵胜利. 面向虚假信息的多模态在线评论情感分析[J]. 河北大学学报(自然科学版), 2025, 45(2): 216-224.

[8] 石洪景. 用户生成内容对消费者跨境电商平台行为意向的影响[J]. 西安电子科技大学学报(社会科学版), 2024, 34(1): 72-81.

[9] 陈远聪. AIGC 在内容生成领域的落地实践与效能提升策略分析[J]. 信息与电脑, 2025, 37(24): 68-70.

[10] 李世勇, 杨铮铮. 基于 BERT-BiLSTM-GAT 的人工智能生成电商虚假评论识别研究[J]. 管理学报, 2025, 22(3): 557-567.

[11] Wu, Y., Ngai, E.W.T., Wu, P. and Wu, C. (2020) Fake Online Reviews: Literature Review, Synthesis, and Directions for Future Research. *Decision Support Systems*, **132**, Article 113280. https://doi.org/10.1016/j.dss.2020.113280

[12] Wang, Y., Pan, Y., Yan, M., Su, Z. and Luan, T.H. (2023) A Survey on ChatGPT: AI-Generated Contents, Challenges, and Solutions. *IEEE Open Journal of the Computer Society*, **4**, 280-302. https://doi.org/10.1109/ojcs.2023.3300321

[13] Amos, C. and Zhang, L. (2024) Consumer Reactions to Perceived Undisclosed ChatGPT Usage in an Online Review Context. *Telematics and Informatics*, **93**, Article 102163. https://doi.org/10.1016/j.tele.2024.102163

[14] Popa, R.G. and Chenic, A.S. (2025) Artificial Intelligence, Consumer Trust and the Promotion of Pro-Environmental Behavior among Youth. *Sustainability*, **17**, Article 5885. https://doi.org/10.3390/su17135885

[15] Cao, C. (2023) The Impact of Fake Reviews of Online Goods on Consumers. *BCP Business & Management*, **39**, 420-425. https://doi.org/10.54691/bcpbm.v39i.4208

[16] Santos, A.M. and Antonio, N. (2025) Improving Trust in Online Reviews: A Machine Learning Approach to Detecting Artificial Intelligence-Generated Reviews. *Information Technology & Tourism*, **27**, 739-766. https://doi.org/10.1007/s40558-025-00329-z

[17] Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., *et al*. (2023) Testing of Detection Tools for AI-Generated Text. *International Journal for Educational Integrity*, **19**, Article No. 26. https://doi.org/10.1007/s40979-023-00146-z

[18] Kovács, B. (2024) The Turing Test of Online Reviews: Can We Tell the Difference between Human-Written and GPT-4-Written Online Reviews? *Marketing Letters*, **35**, 651-666. https://doi.org/10.1007/s11002-024-09729-3

[19] 翟尚铭. 信息生态理论视角下 AIGC 虚假信息治理研究[J/OL]. 河南警察学院学报, 2025, 1-16. https://doi.org/10.16231/j.cnki.jhpc.20251230.001, 2025-12-31.