



The Mediating Role of Trust in the Relationship Between AI-Generated Reviews, Influencer Transparency and Purchase Intentions

Srinivas S¹, Dr. Lakshminarayana K²

¹Assistant Professor, Department of MBA & Research Centre, SJC Institute of Technology, Chikkaballapur, Karnataka, India.

Research Scholar, Department of Management Studies, Visvesvaraya Technological University, Belagavi, Karnataka, India.

Email Id: srinivasvtus scholar@gmail.com, Orcid Id: <https://orcid.org/0009-0005-7463-3585>

²Assistant Professor & Research Supervisor, Department of Management Studies, Visvesvaraya Technological University, Belagavi, Karnataka, India. Email: appinarayan@gmail.com, Orcid Id: <https://orcid.org/0000-0003-0112-3590>

How to Cite this Article:

S, S. (2026). The Mediating Role of Trust in the Relationship Between AI-Generated Reviews, Influencer Transparency and Purchase Intentions. International Journal of Creative and Open Research in Engineering and Management, 2(6).

<https://doi.org/10.55041/ijcope.v2i6.291>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i6.291>

Abstract

The rapid proliferation of AI-generated content in digital commerce has reshaped the dynamics of consumer trust in ways that are not yet fully understood. This study inspects how exposure to AI-generated fake reviews and bot engagement, influencer transparency, and perceived authenticity jointly shape consumer trust, brand credibility, and purchase intentions. Drawing on a quantitative survey of 402 respondents across diverse age groups and product categories, the study employs Pearson's correlation analysis, mediation analysis, moderation analysis (linear regression), exploratory factor analysis, and chi-square tests. Findings reveal a significant "transparency paradox" wherein explicit influencer disclosure weakly but significantly reduces trust, while exposure to AI-generated fake engagement exercises a moderate negative influence on purchase intentions. Perceived authenticity and consumer skepticism emerge as the strongest mediating construct, explaining a extensive portion of the trust–intention relationship. The study contributes to source credibility theory, dual-process cognition, and heuristic persuasion frameworks by demonstrating that consumers simultaneously recognize inauthenticity and respond to volume-based social proof cues. Practical inferences are accessible for brands, content platforms, and regulatory bodies seeking to design ethical and effective digital marketing ecosystems.

Keywords: AI-generated fake reviews, influencer transparency, consumer trust, perceived authenticity, purchase intention, brand credibility, social proof



1. Introduction

Consumer decision-making in digital markets no longer unfolds in a vacuum. With the escalation of social media platforms, influencer marketing has emerged as most powerful vehicles for shaping brand perceptions—and yet, it is also among the most contested. Alongside genuine human influence, a new and troubling ecosystem of AI-generated fake reviews, coordinated bot engagement, and algorithmically amplified social proof has taken root, fundamentally blurring the line between authentic recommendation and manufactured consensus (Filieri et al., 2018; Luo et al., 2020). Regulators, brands, and consumers alike are grappling with the consequences. In India alone, the influencer marketing sector is projected to surpass ₹3,500 crore by 2028 (ASCI, 2023), yet trust deficits are widening as disclosure fatigue and synthetic engagement erode confidence. The Advertising Standards Council of India has strengthened mandatory disclosure norms, yet compliance remains inconsistent. Against this backdrop, a pressing question emerges: does transparency actually build trust, or can it paradoxically undermine it? Prior studies have treated influencer transparency, fake reviews, and consumer trust as largely separate phenomena. What is missing is an integrative empirical examination that captures the complex interplay between these constructs—particularly the mediating role of perceived authenticity and consumer skepticism, and the moderating stimulus of social media usage frequency. This study unswervingly addresses that gap.

Explicitly, this paper scrutinizes how exposure to AI-generated fake reviews and bot engagement (AIFRBE), influencer transparency and ethical perceptions (ITEP), and perceived authenticity, consumer skepticism and social proof (PACSSP) collectively shape consumer trust, purchase intentions and brand credibility (CTPIBC) across a diverse sample of 402 Indian digital consumers. The study makes three principal contributions. First, it empirically surfaces a 'transparency paradox'—the counterintuitive finding that explicit sponsorship disclosure may slightly reduce rather than enhance trust. Second, it demonstrates full mediation of the influencer–trust relationship via PACSSP, highlighting skepticism as a serious intervening mechanism. Third, it shows that social media usage frequency significantly moderates the authenticity–trust relationship, with high-frequency users being more guarded in their trust responses.

Together, these findings enrich source credibility theory, dual-process cognition frameworks, and heuristic persuasion models in the context of AI-mediated digital commerce.

2. Literature Review

2.1 Source Credibility and Consumer Trust

Hovland et al.'s (1953) foundational source trustworthiness model speculates that message persuasiveness rest on on the apparent proficiency and honesty of the communicator. In digital contexts, this model has been extended to encompass influencers and online reviews as critical trust signals (Ohanian, 1990; Cheung & Thadani, 2012). When sources are professed as credible, consumers are further probable to accept their recommendations and proceed to purchase (Erdem & Swait, 2004). However, credibility is a fragile construct—susceptible to manipulation by AI-generated content that mimics authentic voice while lacking genuine human experience (Luo et al., 2020).

2.2 AI-Generated Fake Reviews and Bot Engagement

Fake reviews represent one of the most pervasive forms of online deception. Mayzlin et al. (2014) documented systematic patterns of biased reviews across hotel platforms, demonstrating competitive manipulation. With generative AI, the volume, linguistic quality, and plausibility of fake reviews have increased exponentially (Filieri et al., 2018). Bot-generated engagement—likes, shares, and follower counts—creates artificial social proof that triggers heuristic-based judgments among consumers (Cialdini, 2001). Research steadily shows that consumers who detect or suspect fake reviews report sharply lower trust and brand evaluations (Hu et al., 2012; Luo et al., 2020).

2.3 Influencer Transparency and the Disclosure Paradox

Disclosure regulations—such as the U.S. FTC guidelines (2019) and India's ASCI guidelines (2021)—mandate that influencers label sponsored content. A substantial body of literature suggests that transparency fosters trust by signaling honesty (De Veirman et al., 2017). Yet an emerging counter-narrative documents a 'disclosure backfire effect': when consumers encounter explicit sponsorship labels, they may activate persuasion knowledge and become more resistant to the message (Campbell & Kirmani, 2000; Boerman et al., 2017). Lou and Yuan (2019) find that disclosure can reduce perceived sincerity, particularly when the influencer-brand fit is low. These mixed



findings motivated the present study's hypothesis that transparency may paradoxically weaken trust under certain conditions.

2.4 Perceived Authenticity, Skepticism, and Social Proof

Authenticity is frequently cited as the currency of contemporary influencer marketing (Audrezet et al., 2020). Consumers calibrate perceived authenticity through consistency of content, alignment between espoused values and behavior, and the absence of visible commercial intent (Schallehn et al., 2014). Consumer skepticism—defined as a tendency to disbelieve marketing claims—has grown alongside increased exposure to sponsored content (Obermiller & Spangenberg, 1998). Interestingly, Petty and Cacioppo's (1986) elaboration likelihood model (ELM) predicts that consumers under high elaboration conditions will scrutinize message quality rather than rely on peripheral cues, while those under low elaboration may default to social proof heuristics such as follower counts or review volumes. This dual-processing dynamic is central to understanding why AI-generated engagement can simultaneously raise skepticism and trigger bandwagon effects.

3. Methodology

✓ *Research Design:* A cross-sectional quantitative survey design was adopted. A structured questionnaire was directed online to 402 respondents spanning five age cohorts: below 18 ($n = 39$), 18–24 ($n = 165$), 25–34 ($n = 64$), 35–44 ($n = 86$), and 45 and above ($n = 48$). Purposive and snowball sampling were employed to reach active social media users who had made at least one influencer-prompted purchase within the preceding six months.

✓ *Instrument Development:* Four multi-item constructs were operationalized: AIFRBE (3 items), ITEP (4 items), PACSSP (4 items), and CTPIBC (5 items), all adapted from validated scales in the extant literature. Items were rated on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Exploratory Factor Analysis (EFA) with ProMax rotation confirmed a single dominant factor for AIFRBE, with the construct accounting for 63.4% of cumulative variance (Eigenvalue = 2.198).

✓ *Reliability:* Internal consistency was judged by means of Cronbach's Alpha. CTPIBC demonstrated the highest reliability ($\alpha = 0.840$; 95% CI: 0.786–0.894), followed by AIFRBE ($\alpha = 0.803$; 95% CI: 0.763–

0.843). Both PACSSP ($\alpha = 0.745$) and ITEP ($\alpha = 0.739$) met the acceptable threshold of $\alpha > 0.70$ (Nunnally, 1978). Table 1 presents the reliability summary.

Table 1. Reliability Analysis Summary

Construct	α	95% CI	Interpretation
AI-Generated Fake Reviews & Bot Engagement (AIFRBE)	0.803	0.763–0.843	Good
Influencer Transparency & Ethical Perceptions (ITEP)	0.739	0.636–0.842	Acceptable
Perceived Authenticity, Skepticism & Social Proof (PACSSP)	0.745	0.684–0.806	Acceptable
Consumer Trust, Purchase Intentions & Brand Credibility (CTPIBC)	0.840	0.786–0.894	Good

Source: Primary Data ($n = 402$)

✓ *Analytical Techniques:* Pearson's bivariate correlation was used to test H1–H5. Mediation analysis via Maximum Likelihood (ML) estimation examined whether PACSSP mediates the ITEP–CTPIBC relationship. Moderation analysis through hierarchical linear regression explored whether social media usage frequency moderates the PACSSP–CTPIBC relationship. Chi-square tests assessed age- and gender-based differences in product category engagement.

✓ *Ethical Considerations:* Informed consent was gained from all participants. Anonymity was ensured through de-identification of responses. The study followed institutional ethical guidelines, and no personally identifiable information was retained. All data were deposited steadily and used solely for academic purposes.



4. Objectives and Hypotheses

The study pursues four primary objectives:

- ✓ To examine the direct relationships between AIFRBE, ITEP, PACSSP and CTPIBC
- ✓ To test whether PACSSP mediates the consequence of ITEP on CTPIBC
- ✓ To appraise whether social media usage frequency moderates the PACSSP–CTPIBC relationship
- ✓ To identify age- and gender-based heterogeneity in influencer-driven product category engagement.

H1: There is a noteworthy association between influencer transparency and CT and PI.

H2: There is a noteworthy association between AIFRBE and CTPI.

H3: There is a noteworthy association between AIFRBE and PACSSP.

H4: There is a noteworthy association between IT and PACSSP.

H5: There is a noteworthy association between PA and CTPI.

H6 (Moderation): Social media usage frequency meaningfully moderates the association between PA, CT and PI.

5. Findings and Discussion

5.1 Correlation Analysis (H1–H5)

Table 2. Hypothesis Testing: Pearson's Correlation Summary

H#	Hypothesis	r	P-value	Result
H1	Influencer Transparency → Consumer Trust & Purchase Intention	−0.167 (Weak −ve)	< .001	Supported
H2	AI Fake Reviews & Bot Engagement → Consumer Trust & Purchase Intention	−0.283 (Mod −ve)	< .001	Supported
H3	AI Fake Reviews & Bot Engagement → Perceived	0.294 (Mod +ve)	< .001	Supported

	Authenticity & Skepticism			
H4	Influencer Transparency → Perceived Authenticity & Skepticism	0.195 (Weak +ve)	< .001	Supported
H5	Perceived Authenticity & Skepticism → Consumer Trust & Purchase Intention	−0.456 (Mod–Strong −ve)	< .001	Supported

Source: Primary Data (n = 402)

H1 (ITEP → CTPIBC): A statistically noteworthy but weak negative association ($r = -0.167$, $p < .001$) was found amongst influencer transparency and consumer trust, supporting H1 in an unexpected direction. Rather than confirming that disclosure builds trust, this finding reveals a 'transparency paradox.' When influencers explicitly flag commercial relationships, consumers activate persuasion knowledge (Friestad & Wright, 1994), which can suppress rather than amplify trust. This mirrors the disclosure backfire literature (Boerman et al., 2017) and suggests that formulaic disclosures—absent a genuinely relatable narrative—may register as signals of inauthenticity rather than honesty.

H2 (AIFRBE → CTPIBC): A modest negative correlation ($r = -0.283$, $p < .001$) confirms that exposure to AI-generated fake reviews and bot engagement meaningfully erodes consumer trust and purchase intentions. This is consistent with source credibility theory (Hovland et al., 1953): artificially manufactured engagement diminishes perceived source expertise and trustworthiness. As AI tools make fake reviews ever more linguistically polished, consumer detection mechanisms are put under strain, yet distrust persists at a diffuse level—harming both review platforms and the brands they showcase (Luo et al., 2020).

H3 (AIFRBE → PACSSP): Perhaps the most counterintuitive finding is a moderate positive correlation ($r = 0.294$, $p < .001$) between AI-generated fake engagement and perceived authenticity/social proof. This reflects the dual-processing mechanism described by Petty and Cacioppo (1986): consumers



may intellectually recognize artificially inflated metrics while emotionally responding to them as signals of popularity. Volume of engagement—even when suspected to be artificial—can activate bandwagon heuristics (Cialdini, 2001), leading to a paradoxical elevation of perceived social proof. This underscores the exploitation of cognitive shortcuts in digital environments.

H4 (ITEP → PACSSP): A weak positive association ($r = 0.195$, $p < .001$) suggests that transparent influencer behavior mildly enhances authenticity perceptions. This is directionally consistent with theory (Audrezet et al., 2020), though the small effect size infers that transparency alone is insufficient. Narrative coherence, consistent content style, and lived experience are likely stronger determinants of perceived authenticity—disclosure without these elements may go largely unnoticed in terms of trust formation.

H5 (PACSSP → CTPIBC): The strongest bivariate association in the model ($r = -0.456$, $p < .001$) confirms that PACSSP substantially diminishes consumer trust and purchase intentions. As consumers grow increasingly skeptical of curated, commercially motivated content, their inclination to trust and purchase declines sharply. This finding is pivotal: skepticism—once a peripheral moderating variable—now functions as a powerful mediating force, and brands that fail to address it structurally will face continued trust deficits (Obermiller & Spangenberg, 1998).

5.2 Mediation Analysis

Maximum Likelihood mediation analysis reveals that PACSSP fully mediates the effect of ITEP on CTPIBC. The indirect effect ($\beta = -0.086$, $SE = 0.023$, $z = -3.770$, $p < .001$; 95% CI: -0.130 to -0.041) is statistically significant, while the direct path ITEP → CTPIBC ($\beta = -0.081$, $p = 0.072$) falls below significance once the mediator is included. The total effect remains significant ($\beta = -0.167$, $p < .001$). This complete mediation indicates that the negative impact of influencer transparency on trust operates entirely through heightened consumer skepticism and disrupted authenticity perceptions—not as a direct effect. Brands and influencers must therefore address skepticism at its root rather than assuming that disclosure alone will suffice.

5.3 Moderation Analysis (H6)

Hierarchical linear regression (M_1 : $R^2 = 0.150$, Adjusted $R^2 = 0.137$, $F(6, 395) = 11.609$, $p < .001$) demonstrates that social media usage frequency significantly moderates the PACSSP–CTPIBC relationship. Compared to infrequent users, 'sometimes' users ($\beta = 2.240$, $p < .001$), 'often' users ($\beta = 3.710$, $p < .001$), and 'very frequent' users ($\beta = 2.322$, $p < .001$) show markedly higher baseline trust. However, the interaction terms (PACSSP × $FREQ_Sometimes$: $\beta = -0.501$, $p = 0.002$; PACSSP × $FREQ_Often$: $\beta = -0.819$, $p < .001$; PACSSP × $FREQ_VeryFrequent$: $\beta = -0.463$, $p = 0.004$) confirm that as perceived skepticism rises, the trust-reducing effect is significantly amplified among more frequent users. High-frequency social media users are not naive; they are more attuned to performative authenticity and quicker to penalize brands when their skepticism is triggered.

5.4 Chi-Square Tests: Age and Gender Differences

Age and gender shape influencer-driven engagement in ways that brands cannot afford to overlook. Fashion stood out sharply — younger consumers aged 18–24 dominated this space ($\chi^2 = 30.410$, $p < .001$), while the 45+ group gravitated, perhaps surprisingly, toward Health & Fitness ($\chi^2 = 17.790$, $p = .001$). Home & Kitchen followed a similar age-dependent pattern ($\chi^2 = 15.506$, $p = .004$). Gender told its own story: women leaned into Beauty & Personal Care, men toward Health & Fitness, with Home & Kitchen splitting along predictable lines — all statistically significant.

6. Implications

For Brands and Marketers. Brands must move beyond legal checkbox thinking — instead, building transparency into the story itself, so it reads as honest rather than obligatory. Vetting influencer partners for genuine follower bases is no longer optional; bot-inflated numbers actively damage credibility.

For Influencers and Content Creators. authenticity is not aesthetic, it is architecture. Audiences who scroll daily are sharper, quicker to sense performance, and far less forgiving when trust wavers. Pair every disclosure with real, values-rooted storytelling. Consistency over time builds what a single sponsored post never can: belief.



For Policymakers and Platforms. Regulatory bodies such as ASCI and global equivalents should move beyond binary disclosure requirements toward ecosystem-level interventions—including AI-detection tools for fake reviews, platform-level certification of authentic follower counts, and graduated enforcement for repeat violators. Digital platforms bear responsibility for algorithmic amplification of synthetic engagement and should be held accountable for it.

For Theory. This study extends source credibility theory into the AI-augmented influencer marketing domain and provides empirical support for dual-process cognition in the context of fake engagement. The complete mediation via PACSSP advances a richer understanding of trust formation—one where skepticism is not a peripheral covariate but a central mediating mechanism. The moderation findings integrate frequency-of-use as a substantive boundary condition for authenticity effects.

7. Conclusion

The intersection of artificial intelligence and influencer marketing has created a complex trust environment that defies simple characterization. This study demonstrates, through rigorous empirical analysis of 402 digital consumers, that the relationships between transparency, fake engagement, authenticity, and trust are neither linear nor intuitive. Influencer disclosure does not automatically yield greater trust; AI-generated fake engagement simultaneously erodes trust and can amplify social proof through heuristic channels; and consumer skepticism is the central mediating force that translates influencer behavior into trust and purchase outcomes. What emerges is a image of a consumer population that is simultaneously more skeptical and more susceptible—more aware of manipulation yet still vulnerable to volume-based social cues. For practitioners, this demands authenticity as a long-term strategic investment somewhat than a short-term tactic. For scholars, it calls for continued inquiry into the evolving trust architectures of AI-mediated markets. And for regulators, it highlights the inadequacy of disclosure-only frameworks in an ecosystem where artificial intelligence can generate compelling, deceptive content at industrial scale. Future research should extend these findings through longitudinal designs, platform-specific comparisons, and experimental methods capable of isolating causal

mechanisms. Cross-cultural investigations would also clarify whether the transparency paradox and dual-processing dynamics observed here are specific to the Indian digital context or reflect broader global patterns.

References

- 1) Audrezet, A., de Kerviler, G., & Guidry Moulard, J. (2020). Authenticity under threat: When social media influencers need to go beyond self-presentation. *Journal of Business Research*, 117, 557–569.
<https://doi.org/10.1016/j.jbusres.2018.07.008>
- 2) Boerman, S. C., Willemsen, L. M., & Van Der Aa, E. P. (2017). 'This post is sponsored': Effects of sponsorship disclosure on persuasion knowledge and electronic word of mouth in the context of Facebook. *Journal of Interactive Marketing*, 38, 82–92.
<https://doi.org/10.1016/j.intmar.2016.12.002>
- 3) Campbell, M. C., & Kirmani, A. (2000). Consumers' use of persuasion knowledge: The effects of accessibility and cognitive capacity on perceptions of an influence agent. *Journal of Consumer Research*, 27(1), 69–83.
<https://doi.org/10.1086/314309>
- 4) Cheung, C. M. K., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, 54(1), 461–470.
<https://doi.org/10.1016/j.dss.2012.06.008>
- 5) Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). Allyn & Bacon.
- 6) De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through Instagram influencers: The impact of number of followers and product divergence on brand attitude. *International Journal of Advertising*, 36(5), 798–828.
<https://doi.org/10.1080/02650487.2017.1348035>
- 7) Erdem, T., & Swait, J. (2004). Brand credibility, brand consideration, and choice. *Journal of Consumer Research*, 31(1), 191–198.
<https://doi.org/10.1086/383434>
- 8) Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information & Management*, 55(8), 956–970.
<https://doi.org/10.1016/j.im.2018.04.010>



- 9) Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, 21(1), 1–31. <https://doi.org/10.1086/209380>
- 10) Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change*. Yale University Press.
- 11) Hu, N., Pavlou, P. A., & Zhang, J. (2012). Why do online product reviews have a J-shaped distribution? Overcoming biases in online word-of-mouth communication. *Communications of the ACM*, 52(10), 144–147. <https://doi.org/10.1145/1562764.1562800>
- 12) Lou, C., & Yuan, S. (2019). Influencer marketing: How message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, 19(1), 58–73. <https://doi.org/10.1080/15252019.2018.1533501>
- 13) Luo, X., Gu, B., Zhang, J., & Phang, C. W. (2020). Expert blogs and consumer perceptions of competing brands. *MIS Quarterly*, 44(2), 783–809. <https://doi.org/10.25300/MISQ/2020/13949>
- 14) Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8), 2421–2455. <https://doi.org/10.1257/aer.104.8.2421>
- 15) Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- 16) Obermiller, C., & Spangenberg, E. R. (1998). Development of a scale to measure consumer skepticism toward advertising. *Journal of Consumer Psychology*, 7(2), 159–186. https://doi.org/10.1207/s15327663jcp0702_03
- 17) Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, 19(3), 39–52. <https://doi.org/10.1080/00913367.1990.10673191>
- 18) Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123–205. [https://doi.org/10.1016/S0065-2601\(08\)60214-2](https://doi.org/10.1016/S0065-2601(08)60214-2)
- 19) Schallehn, M., Burmann, C., & Riley, N. (2014). Brand authenticity: Model development and empirical testing. *Journal of Product & Brand Management*, 23(3), 192–199. <https://doi.org/10.1108/JPBM-06-2013-0339>