Proceedings of the 5th International Conference on Interdisciplinary Humanities and Communication Studies
DOI: 10.54254/2753-7064/2026.39177

The Authenticity Deficit and Trust Reconstruction in AIGC-Driven Commercial Communication: Analysis Based on Signaling Theory

Ying Qiao, Xiuling Jia*

Taiyuan University of Technology, Taiyuan, China
**Corresponding Author. Email: jiaxiuling@tyut.edu.cn*

Abstract. Generative artificial intelligence has lowered the barrier to entry for producing high-cost signals in commercial communication. Drawing on signaling theory and combining theoretical deduction with public text analysis, this study examines the mechanism behind the authenticity deficit triggered by AIGC and the limitations of platform labeling systems. The study finds that when generation costs approach zero, cost differences among senders disappear, low-quality senders can easily imitate high-quality signals, and consumers lose the ability to infer genuine effort from content appearance, the distinguishing function of signals collapses, and the market shifts toward a pooling equilibrium. Although mandatory platform labeling provides an institutional alternative, the cost of circumvention is far lower than that of genuine production, making it difficult to unilaterally restore a separation equilibrium. The conclusion indicates that maintaining commercial trust requires a shift from relying on senders' spontaneous cost investments to relying on third-party intervention that substantially curtails the profits derived from violations, and that the focus of regulation should shift from formal review to correcting substantive incentive distortions.

Keywords: AIGC, Signaling Theory, Authenticity Deficit, Institutional Signal, Platform Governance

1. Introduction

The widespread adoption of generative AI in commercial communications has made it possible to mass-produce, at extremely low cost, high-cost signals, such as customer reviews and images of best-selling products, that previously required real-life scenarios and live-action footage. When generating seemingly authentic evidence no longer requires substantial investment, the role of content quality as a distinguishing factor is severely weakened.

On the surface, AIGC appears to enhance the overall quality of content. However, existing empirical research shows that when consumers perceive marketing content as AI-generated, their positive word-of-mouth and trust decline significantly [1, 2]. This paradox cannot be explained by poor content quality but should be attributed to a shift in the information conveyed by the act of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7064/2026.39177

generation itself. When high-quality content becomes readily accessible to everyone at low cost, consumers lose the ability to infer genuine effort based on the appearance of content.

This requires explanation from the perspective of signaling theory. The effectiveness of credible signals depends on observable cost differences among senders. As He et al. point out, large language models have made fluent writing cheap, shifting the original separation equilibrium toward a pooling equilibrium [3]. When the cost of generating high-quality content approaches zero, low-quality content is no longer excluded by cost barriers, and signals lose their discriminatory function. This is precisely the mechanistic root cause of the loss of authenticity in commercial communication.

Centering on the mechanism of signal cost collapse, this study analyses the formation logic of the authenticity deficit in AIGC-driven commercial communication and examines the intervention limits of platform AI labeling systems as an institutional signal. The analytical materials are primarily drawn from publicly available regulatory cases and consumer survey data.

2. Literature review

Regarding the impact of AIGC on commercial communication, existing research has accumulated rich empirical findings, but most discussions focus on how consumers evaluate AI-generated content or whether disclosure mechanisms are effective. Relatively few explain the root of the trust crisis from the perspective of the cost structure of sender groups. This section reviews three main research strands and their limitations, and then introduces signaling theory to explain what kind of disruption AIGC has caused to this theoretical premise.

2.1. Existing stands and limitations of research on AIGC commercial communication

The first strand focuses on consumers' perceived authenticity of AIGC. Experimental evidence indicates that even when AI-generated visual and textual content is visually indistinguishable from human creations, consumers still systematically evaluate it as insufficiently authentic [1]. In the context of virtual influencers, perceived authenticity has been confirmed as a key mediating variable linking AI use and consumer trust, with transparency significantly moderating this process [4]. This strand confirms that AIGC has indeed altered consumers' judgment processes, but its unit of analysis is confined to the individual psychological mechanism from cue to perception, failing to explain why the same cue triggers a different weight under AIGC conditions than in the past.

The second strand focuses on the effects of AI disclosure strategies. Koning and Voorveld's experiment found that AI disclosure simultaneously activates consumers' conceptual knowledge and persuasive knowledge, the former may promote transparency, while the latter leads to a decline in trust in advertisements and brands [2]. A series of experiments by To et al. in the context of luxury brands further revealed that disclosing AI-generated imagery leads consumers to infer that the brand has invested less effort, thereby reducing advertising authenticity and brand evaluation [5]. The value of this strand lies in revealing the implicit effort inference logic in consumer judgment. However, effort is usually measured as a mediating variable at the individual psychological level and has not been elevated to the level of cost differences among sender groups.

The third strand adopts an institutional and governance orientation, discussing how tools such as platform labeling and regulatory disclosure address the information problems of AIGC. Related studies distinguish between declarative, regulatory, and infrastructural forms of disclosure, and point out that different forms shift attribution from the brand to the platform or technological system [6]. This strand shifts attention from individual consumers to institutional arrangements. However, its

DOI: 10.54254/2753-7064/2026.39177

analytical framework still centers on whether disclosure is sufficient, and it rarely explores whether, under the condition that signal costs have already collapsed, the labeling system is repairing the distinguishing function of signals or merely institutionalizing the fact of signal failure. Together, the three strands point to a gap that has not been fully theorized: the impact of AIGC on trust is rooted in the structural disappearance of cost differences among senders.

2.2. Core conditions of signaling theory: cost differences and separation equilibrium

Spence's labor market signaling model is based on a strict condition: signals convey information precisely because the costs of sending the same signal differ among senders of different types [7]. In a separation equilibrium, low-quality senders find it unprofitable to imitate high-cost signals and therefore choose not to send them, making the signal a reliable indicator of type. Akerlof's lemon market logic further demonstrates that when buyers cannot distinguish quality before purchase, low-quality goods systematically crowd out high-quality ones [8]. Together, these two classic models illustrate a premise: information asymmetry can be mitigated by signals because the cost of signals is asymmetric for different types of senders.

When this framework is transposed to commercial communication, the credibility of high-cost signals in traditional commercial communication, such as real-person live-action footage, detailed reviews requiring genuine consumption experience, and brand content supported by physical settings, does not stem from consumers' belief that aesthetic appeal equates to authenticity, but rather from the fact that producers have invested time, settings, and performance risks in sending these signals that cannot be easily replicated by low-quality types. The cost differences of signals constitute the implicit foundation for consumers' inferences.

2.3. Asymmetric impacts of AIGC on the cost structure of signals

AIGC does not alter the appearance of signals, but rather the distribution of marginal costs associated with producing them. The synthetic sincerity model proposed in the field of organizational behavior clearly illustrates this mechanism. Generative AI reduces the marginal production costs of empathetic and fluent language, causing the inferred sending costs of receivers to decline accordingly, which in turn weakens perceived sincerity and interpersonal trust. A study summarizes this process as the cost- collapse pathway, wherein increasing reliance on GenAI leads to lower inferred costs by receivers, subsequently triggering a decline in perceived sincerity and a decay of trust [9]. The equilibrium analysis by He et al. further points out that large language models make fluent writing cheap, causing the equilibrium of content signals to shift from separating toward semi-separating or even pooling, resulting in a systematic decay in the credibility of production-side signals [1].

The commercial implication of this shock is that when content appearing as authentic evidence no longer requires the production costs of authentic evidence, low-quality senders are no longer excluded from imitation behavior due to an inability to bear signal costs. High-quality senders may still choose to invest costs, but their investment no longer generates a separating effect, because consumers cannot infer from the signal itself whether the cost has been paid. Signals degrade from indicators of cost differences to cost-free appearance replications, and their distinguishing function collapses accordingly. It must be emphasized that this inference does not imply that all AIGC content is untrustworthy. Rather, it points to a more specific proposition that when the cost conditions of production-side signals are systematically undermined, the spontaneous signaling mechanism relying on cost differences to transmit information will be unable to maintain a

DOI: 10.54254/2753-7064/2026.39177

separating equilibrium. The resulting theoretical demand is whether an alternative signaling mechanism exists that does not depend on the production cost differences of senders. This is precisely the core question when analysing the platform labeling system as an institutional signal in subsequent chapters.

3. Research methods

3.1. Research design and analytical framework

This study adopts a strategy combining theoretical deduction with public texts analysis. Since the core objective of this study is to explain how signaling mechanisms fail when the cost structure undergoes a fundamental change and to reveal the underlying causal mechanism and equilibrium shifting process, case and text analysis can provide direct mechanistic explanatory power for this theoretical issue [10]. In practical terms, this study first deduces, based on classical theoretical models, how cost differences among senders constitute the fundamental condition for signaling effectiveness. Subsequently, this deduction result is compared with regulatory cases and platform governance practices that have already occurred in reality, thereby reconstructing the failure process of the consumer judgment mechanism. This analytical approach helps identify the mode of action of the mechanism in the interaction between institutional contexts and specific behaviors.

3.2. Materials collection and selection

The selection of analytical materials follows the principles of traceability and representativeness [11]. This study primarily draws on three types of public texts. The first type is public penalty cases from regulatory authorities, such materials directly present typical behaviors of using AIGC to construct false commercial signals and the basis for their determination. The second type consists of AIGC content governance announcements and labeling specifications issued by mainstream platforms, which reflect how platforms, as third parties, attempt to intervene in and reconstruct commercial trust signals. The third type comprises consumer perception and trust survey reports released by authoritative research institutions, these data corroborate the specific manifestations of signal failure in reality from the receiver's side. These three types of texts corroborate each other and collectively form the empirical basis for verifying the signal cost collapse mechanism. In the specific analysis, this study will focus on how these cases and rule texts reflect the disappearance of cost differences among senders, and why the intervention of institutional signals finds it difficult to fully restore the original separating equilibrium.

4. Mechanistic analysis of signal cost collapse and authenticity deficit

4.1. Scaled production of fake signals

In traditional commercial communication, the reason why hot-selling scenes and product usage effects possess signal value is that presenting these scenes requires senders to actually organize supply or undergo a genuine consumption process, the production cost itself constitutes a threshold that is difficult for low-quality senders to cross. AIGC has dismantled this threshold. In the first AI false advertising case investigated by the Liangjiang New Area Market Supervision Bureau in Chongqing, a dealer used AI to synthesize scenes of consumers queuing to make purchases and disseminated them with exaggerated rhetoric, while there was no record of any genuine transaction in the video. Law enforcement authorities characterized the case as entirely AI-generated and purely

DOI: 10.54254/2753-7064/2026.39177

fictional in scene, with a concealment significantly stronger than that of traditional false advertising. Traditional false advertising at least requires genuine shooting activities, locations, and personnel, although these costs cannot guarantee authenticity, they constitute a certain degree of entry threshold. When synthesizing footage that appears to contain genuine transactions no longer requires any investment related to the transaction, a systematic decoupling occurs between the production cost of the signal and the fact it claims to represent.

The extent of this change can be corroborated by platform governance data. In 2025, internet advertising cases accounted for 82% of the total number of advertising violation cases in China, and typical situations investigated by market supervision departments included the use of AI to forge images for publishing advertisements. Governance results disclosed by Ocean Engine during the same period showed that the platform shut down more than 50000 accounts engaged in violating marketing. When the production cost of such content approaches zero and can be iterated in batches, the problem has exceeded the opportunistic behavior of individual senders and points to a systematic lifting of cost constraints in the signal production process.

4.2. Screening failure and authenticity deficit

The direct consequence of signal production cost collapse is not that consumers stop trusting, but that consumers lose a reliable basis for inference based on the appearance of signals. When exquisite buyer reviews can originate from either genuine consumption or zero-cost generation, the quality of the content itself can no longer indicate whether the sender has borne the cost. Consumer uncertainty shifts from evaluating signal quality to evaluating whether the signal corresponds to authentic behavior. Existing survey data corroborates this judgment dilemma. A survey of 1539 U.S. consumers by Gartner in 2025 showed that 68% of respondents frequently doubt the authenticity of the content they see, and consumer behavior is being forced to shift from intuitive judgment to active verification. A 2026 survey by Cint, on the other hand, showed that 36% of respondents described AI-created advertisements as untrustworthy [12]. These data indicate that when consumers realize that the signal cost conditions have failed, their overall evaluation of signal credibility will be adjusted downward.

Experimental research further reveals the micro-mechanism of this downward evaluation adjustment. A study involving 940 participants found that AI-generated advertisements are not more effective the more realistic they are, perceived trust plays a mediating role in the matching between visual style and delivery methods [13]. Another experiment indicated that AI disclosure simultaneously activates consumers' conceptual knowledge and persuasive knowledge, with the latter directly leading to a decline in trust in advertisements and organizations [2]. These findings point to the same conclusion: once consumers perceive that senders have circumvented traditional production costs, they will initiate defensive inference. What is damaged is not only the specific quality of a certain type of content, but also the efficacy of the signal as a distinguishing tool itself.

4.3. Intervention and limits of institutional signals

When cost differences among senders can no longer spontaneously generate a separating equilibrium, alternative signals need to be provided by third parties other than the senders. In March 2025, four departments including the Cyberspace Administration of China jointly issued the Measures for Labeling AI-Generated Synthetic Content, requiring AI-generated text, images, and videos to be clearly labeled and embedded with digital watermarks, effective September of that year. This institutional arrangement attempts to transform whether to disclose AI usage into a new

DOI: 10.54254/2753-7064/2026.39177

observable signal, that is, disclosure does not depend on the production cost of the sender, but is a compliance requirement imposed by rules. Article 50 of the EU Artificial Intelligence Act similarly requires deployers of deep-synthesized content to disclose its artificially generated nature. These arrangements constitute institutional signals independent of the spontaneous behavior of senders.

However, the effectiveness of institutional signals is constrained by an inherent weakness: their sending cost can likewise be lowered. The labeling system relies on proactive declaration or technical identification, while the cost of evading labeling is far lower than that of evading genuine production. The aforementioned Chongqing case, in which the party involved only attached partial topic tags without completing standardized labeling, is a direct example. When the expected benefit of non-labeling is higher than the probability of detection multiplied by the penalty severity, low-quality senders still have the incentive to choose concealment. Institutional signals can narrow the scope of pooling equilibrium but are unlikely to unilaterally rebuild a separating equilibrium unless compliance costs are raised high enough to replace the screening function originally borne by production costs.

5. Discussion

5.1. The transformation of trust production modes

In the traditional commercial environment, senders prove their integrity to the market by bearing high preparation costs, and this spontaneous order maintains the basic trust between trading parties. As the threshold for content preparation universally declines, it is no longer possible to infer the actual investment of senders solely based on appearance quality, and the signaling mechanism at the individual level tends to fail. Trust thus gradually detaches from the spontaneous behavior of senders and transforms into a public resource that requires continuous supply from external institutions. As the dominant actor in the commercial information ecosystem, the role of platforms is shifting from a mere content distribution channel to an underlying guarantor of the trust mechanism. This transformation means that maintaining market trust is no longer merely a game between buyers and sellers, but depends on the depth of intervention by platform rules in information asymmetry.

5.2. Implications for platform governance and advertising regulation

Current governance measures, which primarily rely on label prompts, attempt to alleviate information opacity through mandatory disclosure. However, if the cost of violation is insufficient to cover the cost of evasion, formal labeling will only degenerate into a superficial project of perfunctory compliance. The focus of regulation should shift from formal review to the correction of substantive incentive distortions, for example, linking the generation method of content with algorithmic recommendation weights, or imposing substantive traffic penalties on entities that conceal the fact of generation [14]. Only by making the expected loss of non-compliant behavior far exceed its potential gain can institutional signals truly replace the original production costs and re-establish market thresholds with distinguishing capabilities.

6. Conclusion

This study examines the impact of generative artificial intelligence on the trust mechanism in commercial communication and its institutional responses. The analysis indicates that the large-scale application of AIGC has changed the cost structure of information transmission. When the marginal

DOI: 10.54254/2753-7064/2026.39177

cost of generating realistic content approaches zero, the entry threshold originally maintained by high production costs is dismantled. Low-quality senders can easily imitate high-quality signals, causing market signals to slide from a separating equilibrium toward a pooling equilibrium and triggering a consumer authenticity deficit regarding commercial information. Faced with the failure of spontaneous order, the platform labeling system attempts to provide alternative institutional guarantees through mandatory disclosure. However, since the cost of evading labels is far lower than that of evading genuine production, simple label regulation finds it difficult to unilaterally rebuild the original screening function. The maintenance of trust is shifting from relying on the production costs spontaneously invested by senders to relying on third-party substantive intervention in illicit gains. The theoretical deduction and public text analysis adopted in this study help explain the generation mechanism of the trust crisis at a macro level. Limited by the types of materials, this study did not collect micro-level data at the consumer level and cannot precisely measure the actual effect of institutional signals of different intensities on individual trust repair. Future research can test the differentiated impacts of disclosure strategies and penalty intensities on consumer judgment through experimental designs or questionnaire surveys. With the continuous iteration of generative technology, subsequent research can also focus on the further changes in the signal recognition mechanism caused by multimodal synthetic content, as well as the potential role of cross-platform collaborative governance in reconstructing digital trust.

References

[1] Brüns, J. D., & Meißner, M. (2024). Do you create your content yourself? Using generative artificial intelligence for social media content creation diminishes perceived brand authenticity. *Journal of Retailing and Consumer Services, 79*, 103790. https://doi.org/10.1016/j.jretconser.2024.103790.

[2] Koning, B., & Voorveld, H. A. M. (2025). Disclaimer! This content is AI-generated: How AI-disclosures influence trust in advertisements and organizations. *Journal of Interactive Advertising, 25*(3), 240–253. https://doi.org/10.1080/15252019.2025.2554149.

[3] He, M., Jiang, W., Shaheen, M., Xie, H., Sun, M., Rathore, R. S., & Gu, W. (2026). When content signals lose credibility: Signalling equilibria on human-centred platforms in the LLM era. *Expert Systems, 43*(10), e70403. https://doi.org/10.1111/exsy.70403.

[4] Bui, H. T., Filimonau, V., & Sezerel, H. (2024). AI-thenticity: Exploring the effect of perceived authenticity of AI-generated visual content on tourist patronage intentions. *Journal of Destination Marketing & Management, 34*, 100956. https://doi.org/10.1016/j.jdmm.2024.100956.

[5] To, R. N., Wu, Y. C., Kianian, P., & Zhang, Z. (2025). When AI doesn't sell Prada: Why using AI-generated advertisements backfires for luxury brands. *Journal of Advertising Research, 65*(2), 202–236. https://doi.org/10.1080/00218499.2025.2454120.

[6] Lan, G., Abani, A., Oscar, C., Margaux, R., Zayna, C., Nick, F., Chenhao, T., Kurt, T., & Marshini, C. (2026). Governance of AI-generated content: A case study on social media platforms. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 1604, 1–22. https://doi.org/10.1145/3772318.3790415.

[7] Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics, 87*(3), 355–374. https://doi.org/10.2307/1882010.

[8] Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics, 84*(3), 488–500. https://doi.org/10.2307/1879431.

[9] Marian, H. (2026). Synthetic sincerity: When generative AI makes signals of care cheap at work. *Academy of Management Proceedings, 2026*(1). https://doi.org/10.5465/AMPROC.2026.21811poster.

[10] Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage Publications.

[11] Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027.

[12] Cint. (2026). Consumer attitudes towards AI in advertising. *Cint Research Report*. https://www.cint.com.

[13] Meilian, Q., Kan, J., & Xiaoning, B. (2026). Falsity images also have real benefits? The impact of AI-generated advertisements' visual styles and reach methods on consumer consultation intentions. *Journal of Retailing and*

DOI: 10.54254/2753-7064/2026.39177

Consumer Services, 91(2026), 104772. https://doi.org/10.1016/j.jretconser.2026.104772.
[14] Helberger, N. (2020). The political power of platforms: How current attempts to regulate misinformation amplify opinion power. *Digital Journalism, 8*(6), 842–854. https://doi.org/10.1080/21670811.2020.1773888.