Proceedings of the 6th International Conference on Educational Innovation and Psychological Insights
DOI: 10.54254/2753-7048/2025.26637

How Does Deep Synthesis Content Erode Trust in Social Media? — An Analysis from the Dimensions of Platform, Information, and Users Based on a Survey Experiment

Xinhui Wu[1], Sheng Lin[1*]

[1]School of Law, Hangzhou Dianzi University, Hangzhou, China
**Corresponding Author. Email: l1332286868@163.com*

Abstract. Deep-synthesis technology—an artificial intelligence technique used primarily for audiovisual content—has been misused in ways that are systematically deconstructing the trust ecosystem of social media. This study innovatively unpacks social media trust into three dimensions (platform trust, information trust, and user trust) and, using a survey-experiment design (n = 522), examines the differentiated erosive effects of deep-synthesis content within social media platforms. Taking Douyin (the Chinese version of TikTok) as the context, we employ five categories of deep-synthesis videos as stimuli and control for topic-related confounds. The findings show that deep-synthesis content exerts a negative impact on social media trust, with information trust being the most severely damaged, followed by platform trust; user/interpersonal trust exhibits a lagging and comparatively weak effect. These results reveal differentiated pathways through which technological alienation disrupts trust mechanisms and provide a theoretical basis for platform governance and user-level cognitive interventions.

Keywords: deep-synthesis technology, social media trust, survey experiment

1. Introduction

With the rapid development of AI technologies, deep-synthesis techniques are increasingly permeating social media platforms such as Douyin and Xiaohongshu, presenting content to users in diverse forms. This technology is a multimodal media form generated through deep learning, which can be classified into four categories: images, audio, video, and audio/video synthesis [1]. Its underlying principle lies in the application of deep learning models such as Generative Adversarial Networks (GANs). Within large datasets, the generator is responsible for producing images or audio, while the discriminator distinguishes between synthetic and authentic content; through adversarial training, this process enables highly realistic manipulation or generation of audio-visual content [2].

As a core element of social capital in the digital era, social media trust is facing severe challenges posed by deep-synthesis technologies. By generating realistic audiovisual content, these technologies blur the boundary between reality and fabrication, fundamentally undermining the cognitive foundations on which users' trust in the social media ecosystem rests. Deep-synthesis techniques increase the deceptiveness of online content. According to the 2024 Artificial Intelligence

© 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: 10.54254/2753-7048/2025.26637

Security Report, AI-based deepfake fraud surged by 3000% [3]. In social media environments where deep-synthesis technologies are widely applied, existing research has mainly focused on technological risks (e.g., the spread of false news, infringement of portrait rights) or on single dimensions of trust, while empirical examinations of the structural deconstruction of trust and its differentiated erosion pathways remain scarce.

This study goes beyond the traditional "content–institution" binary framework of media trust and deconstructs social media trust into three dimensions: platform trust (based on systemic mechanisms), information trust (based on content authenticity), and interpersonal trust (based on user-to-user relationships). Using a survey experiment with 522 university students and employing synthetic videos as stimulus materials, this research empirically demonstrates, for the first time, the differentiated erosive effects of deep-synthesis content on the three types of trust. The findings provide targeted insights for addressing the ethical risks of such technologies.

2. Literature review and research hypotheses

2.1. Factors influencing trust in social media

Information is one of the key factors influencing trust in social media. The widespread dissemination of false information on social media disrupts users' normal information behaviors, severely endangering their health and financial security, and giving rise to major crises of trust in social media [4]. The low quality of health-related information on social media makes it difficult for the public to correctly identify reliable health knowledge, thereby undermining social media trust [5]. The platform itself is also an important factor. The effectiveness with which social media corrects misinformation and its patterns of information handling influence users' trust in the platform [6]. A lack of source-verification mechanisms and the diversity of information forms have gradually turned social media into a breeding ground for a "fog of health information," which leads people to frequently avoid social media content altogether [7]. From the perspective of users themselves, strong perceptions of potential risks—such as cognitive overload and privacy breaches—can trigger anxiety, fear, and annoyance, which in turn affect their health-related information behaviors and produce trust crises [8]. At the same time, Chu Yanbo and others have pointed out that users' own level of involvement with social media information has a positive effect on their trust [9].

2.2. Deep-synthesis technology and social media trust

The application of deep-synthesis technology on social media has attracted particular attention because of its deceptive nature. Many scholars have focused on its societal impacts, especially in the field of journalism and media, such as the distortion of political discourse [10] and violations of portrait rights [11]. The uncontrollable spread of fake news within social networks, coupled with information overload and the erosion of trust and authority in traditional news channels, has contributed to the dangerous erosion of democratic values. The introduction of generative technologies into information production and circulation systems may act as a disruptive factor, reducing the perceived credibility of information among social actors [12]. The uncertainty generated in this context may further reduce trust in social media, leading to widespread uncertainty and cynicism, and exacerbating the challenges to online civic culture in democratic societies [13]. Deep-synthesis technology is also associated with the decline of news authenticity, the deepening of the post-truth condition, manipulation of public opinion, disruption of information order, technological alienation, and escalating social trust risks [14].

DOI: 10.54254/2753-7048/2025.26637

2.3. Research hypotheses

Social media trust is a subset of media trust, but it possesses distinctive features while also exhibiting path dependence on prior research in media trust. Traditionally, media trust has been approached from two perspectives: content trust and institutional trust [15]. With respect to social media, Dwyer et al. have divided trust into two types: trust in the social networking site or medium, and trust in other people within social media [16]. In other words, trust on social media is generally a combination of interpersonal trust and system trust. Specifically, interpersonal trust is grounded in interactions and relationships among users, built upon behavioral consistency, honesty, and mutual understanding. System trust, in contrast, relies on the mechanisms and rules of the platform itself. Users' trust in social media platforms is often based on factors such as information security, privacy protection, and content moderation mechanisms [17].

Overall, the academic community has yet to reach consensus on an operationalized definition or standardized measurement tools for social media trust. There remains debate over whether media trust refers to trust in news media in general, trust in different types of media, trust in specific news outlets, or trust in the news content itself [18]. Although in theory media trust may be considered multidimensional, empirical research has not consistently supported this view. Most studies have examined public trust in news media primarily from the perspective of credibility [19]. Following this line of thought, the present study analyzes the uncertainty introduced by deep-synthesis technology into social media and examines its impact on trust across three dimensions: platform trust, source (publisher) trust, and information trust.

Carnevale has noted that the introduction of deep-synthesis technology into social media may serve as a disruptive factor, reducing users' trust in information [12]. Given its inherently deceptive nature, we hypothesize that deep-synthesis technology has negative effects across all three dimensions of social media trust.

H1: The application of deep-synthesis technology in disseminated content negatively affects platform trust.

H2: The application of deep-synthesis technology in disseminated content negatively affects information trust.

H3: The application of deep-synthesis technology in disseminated content negatively affects interpersonal trust.

3. Research method

3.1. Survey–experiment method

This study employs a survey–experiment method, which combines the strengths of both surveys and experiments. The approach has relatively low requirements for experimental settings, is highly operable, and at the same time ensures research validity. By using randomization, the survey–experiment design eliminates systematic differences between the experimental and control groups, thereby addressing endogeneity issues and ensuring the accuracy of results [20]. In this study's experimental design, to avoid self-selection bias and endogeneity problems, a situational experiment was adopted. The application of deep-synthesis technology in disseminated content was randomly manipulated, and participants were placed in a simulated scenario for intervention. Specifically, this within-subject experiment examined how university students' trust in social media changed under exposure to deepfake application scenarios.

DOI: 10.54254/2753-7048/2025.26637

3.2. Experimental design

The experimental procedure included: a pretest of participants' demographic information, a pretest measurement of social media trust, random assignment to view different videos, and finally, a posttest questionnaire measuring social media trust.

Currently, social media is saturated with AI-generated short videos, and it is nearly impossible to find a platform completely free from them. Hence, constructing an experimental environment was necessary to generate control conditions, i.e., to build the moderating variable. In this study, the simulated experimental scenario was modeled on users' routine browsing of Douyin videos. Intervention was achieved by embedding deep-synthesis stimuli within the questionnaire design to construct virtual conditions. The experimental group was exposed to videos containing deep-synthesis technology, while the control group watched ordinary videos. Two sets of five videos were selected. The themes included: singer performances, giant figures, marine creatures, political figures, and a video without deep-synthesis application. In the experimental group, the first four videos employed deep-synthesis techniques. In the control group, all but the political figure video were free of deep-synthesis technology. This design controlled for thematic effects and enhanced internal validity. Moreover, the between-subjects setup enabled evaluation of whether exposure to experimental vs. control videos influenced participants' evaluations of the videos and their trust in social media.

3.3. Experimental sample

Data were collected through multi-wave surveys between October 2024 and December 2024, with the sample consisting of university students in China. The survey was conducted in two rounds, each divided into two stages: the first stage gathered demographic information (e.g., gender, age, education level), and the second stage conducted the experiment. A total of 627 questionnaires were distributed, and 522 valid responses were collected, yielding an effective response rate of 83.2%. From the demographic variables: 51.53% were male and 48.47% female. In terms of education level, 24.33% were junior college students, 62.45% undergraduates, and 13.22% postgraduates or above. Regarding age distribution, the majority of participants were university students aged 18–23, accounting for 71.84% of the sample.

4. Empirical results

4.1. Reliability and validity test

This study employed SPSSAU to conduct reliability analysis. Using Cronbach's α coefficient to evaluate the reliability of the questionnaire, the results show that the overall reliability coefficient was 0.771. The Cronbach's α values of all six variables were greater than 0.6, indicating that the questionnaire used in this study had acceptable reliability.

4.2. Independent-sample t-test

An independent-sample T-test was conducted to examine whether the experimental manipulation of the independent variable was effective in the questionnaire design. The results (see Figure 1 and Table 1) indicate that, under experimental manipulation, the mean values of platform trust, information trust, and interpersonal trust differed between the experimental and control groups. The non-significant differences in pre-test results suggest that group assignment was random. By

DOI: 10.54254/2753-7048/2025.26637

contrast, the significant differences in post-test results indicate that the manipulation of the experimental variable in the questionnaire was effective.

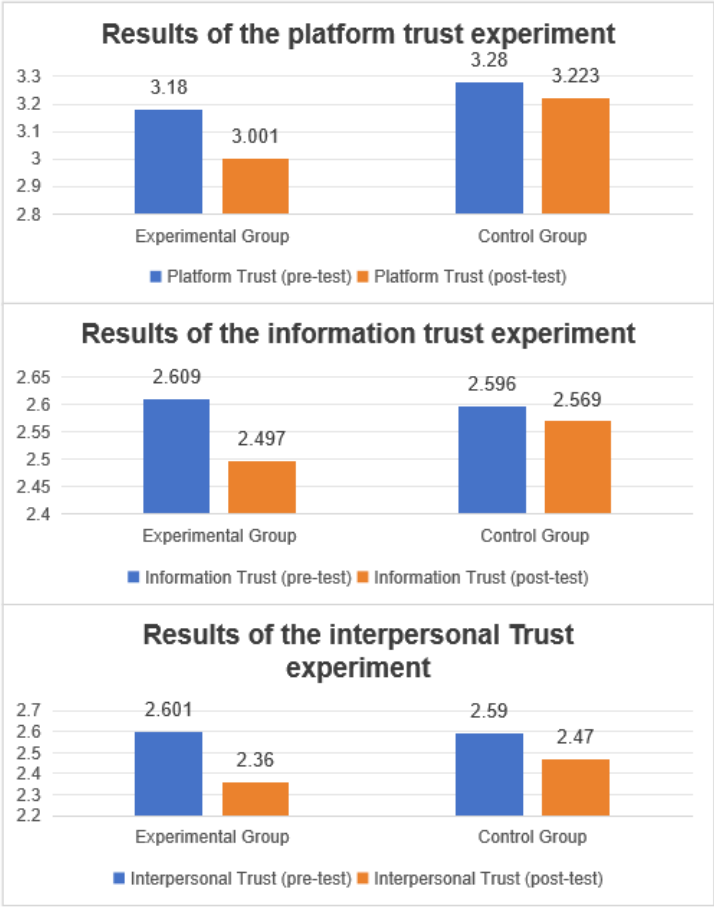


Figure 1. Experimental results of trust across dimensions

Table 1. Independent-sample T-Test

Variable	Assumption	F	t	Sig. (2-tailed)	Mean Difference (Pre–Post)
Platform Trust (Pre)	Equal variances assumed	1.157	-1.213	0.226	-0.098
	Equal variances not assumed		-1.213	0.226	-0.098
Information Trust (Pre)	Equal variances assumed	1.084	0.649	0.517	0.0406
	Equal variances not assumed		0.649	0.517	0.0406
Interpersonal Trust (Pre)	Equal variances assumed	0.063	1.033	0.302	0.096
	Equal variances not assumed		1.033	0.302	0.096
Platform Trust (Post)	Equal variances assumed	0.009	-1.82	0.049*	-0.1223
	Equal variances not assumed		-1.82	0.049*	-0.1223
Information Trust (Post)	Equal variances assumed	1.307	-3.334	0.001**	-0.1994
	Equal variances not assumed		-3.335	0.001**	-0.1994
Interpersonal Trust (Post)	Equal variances assumed	0.122	-1.451	0.014*	-0.117
	Equal variances not assumed		-1.451	0.014*	-0.117

DOI: 10.54254/2753-7048/2025.26637

The results demonstrate that the application of deep-synthesis technology in disseminated content exerts negative effects on all three dimensions of social media trust. Among them, the strongest effect was on information trust (–0.1994), followed by platform trust (–0.1223) and interpersonal trust (–0.117). Therefore, hypotheses H1, H2, and H3 are all supported.

5. Discussion

5.1. Differential erosion mechanisms of three-dimensional trust

This study confirms the overall negative impact of deep-synthesis content on social media trust (H1–H3 supported). However, the intensity and pathways of erosion vary significantly across the three dimensions:

Information trust suffers the most severe damage. Synthetic content forces users to invest additional cognitive resources to distinguish authenticity (e.g., detecting micro-expression anomalies in political leader videos). This leads to cognitive overload, directly undermining confidence in judging the truthfulness of information. Platform trust ranks second. Users tend to attribute the proliferation of synthetic content to the platform's failure in content moderation mechanisms. Participants in the experimental group rated Douyin's "capacity to control false content" significantly lower, corroborating theories that system trust relies heavily on institutional safeguards. Interpersonal trust is the weakest and most delayed in effect. This may stem from the "trust-buffer effect" within social networks. Users are more inclined to doubt unfamiliar sources (e.g., influencer-generated videos) rather than close social circles. This finding supports Chuyambo's argument that "involvement level moderates trust," but it also signals the risk that long-term exposure to synthetic content could erode the very foundations of relational trust.

5.2. Theoretical contributions and practical implications

From a theoretical perspective, this study identifies cognitive load as a key mediating variable, filling a gap in prior research that focused mainly on technological features while neglecting psychological mechanisms. It also confirms the heterogeneous responses across trust dimensions, challenging the assumption that "social media trust is a unitary construct" and refining Dwyer's binary framework into a more nuanced three-dimensional model.

From a practical perspective: Platforms should optimize synthetic content labeling systems (e.g., mandatory watermarks) to reduce users' cognitive burden. Regulatory agencies should prioritize targeting highly deceptive synthetic content (e.g., political videos), as these pose the greatest threat to information trust. User education should emphasize cognitive skill training (e.g., recognizing textural artifacts in AI-generated marine life videos) to help bridge the digital literacy divide and strengthen resilience against synthetic deception.

References

[1] Twomey, J., Ching, D., Aylett, M. P., et al. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. PLOS ONE, 18(10), e291668.
[2] Mirsky, Y., & Lee, W. (2022). The creation and detection of deepfakes. ACM Computing Surveys, 54(1), 1–41.
[3] Qi An Xin Group. (2024). 2024 Artificial Intelligence Security Report [Report]. Qi An Xin Group.
[4] Li, J., Wen, J., Xu, Q., et al. (2024). Continued use or negative use: A study on dynamic user behavior of social media from the perspective of ambivalent attitudes. Information Science, 42(9), 100–111.
[5] Li, Y., Zhang, X., & Wang, S. (2018). Research on the quality of health information in social media: An analysis based on features of true and false health information. Journal of the China Society for Scientific and Technical

DOI: 10.54254/2753-7048/2025.26637

Information, 37(3), 294–304.
[6] Yu, M., Yu, S., & Liu, R. (2024). Healthcare workers' correction intentions of false health information on social media: Based on SEM and fsQCA methods. Journal of Information Resources Management, 14(3), 104–120.
[7] Peng, L., & Jiang, X. (2022). The generative mechanism of health information avoidance behavior of social media users from the perspective of risk perception. Library and Information Service, 66(22), 55–65.
[8] Gallotti, R., Valle, F., Castaldo, N., et al. (2020). Assessing the risks of "infodemics" in response to COVID-19 epidemics. Nature Human Behaviour, 4(12), 1285–1293.
[9] Chu, Y., Wang, P., & Hu, S. (2025). The mechanism of social media health information user trust formation. Modern Information, 45(1), 97–111.
[10] Zhang, A., & Wang, F. (2021). Deepfakes and the mutation of political opinion from the perspective of artificial intelligence. Journal of Hohai University (Philosophy and Social Sciences), 23(4), 29–36.
[11] Karasavva, V., & NoorbhaI, A. (2021). The real threat of deepfake pornography: A review of Canadian policy. Cyberpsychology, Behavior, and Social Networking, 24(3), 203–209.
[12] Carnevale, A., Delgado, C. F., & Bisconti, P. (2023). Hybrid ethics for generative AI: Some philosophical inquiries on GANs. Humana. Mente Journal of Philosophical Studies, 16(44), 33–56.
[13] Leyva, R., & Beckett, C. (2020). Testing and unpacking the effects of digital fake news: On presidential candidate evaluations and voter support. AI & Society, 35, 969–980.
[14] Zhang, D. (2023). The impact and governance of deepfake technology on journalism. Youth Journalist, (23), 41–45.
[15] Zhou, Z. (2024). Rediscovering the state: An attempt to expand the concept of media trust in China. Chinese Journal of Journalism & Communication, 46(7), 28–53.
[16] Dwyer, C., Hiltz, S., & Passerini, K. (2007). Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. AMCIS 2007 Proceedings, 339.
[17] Niu, J., & Meng, X. (2019). The influence of social media trust on privacy risk perception and self-disclosure: The mediating effect of online interpersonal trust. Chinese Journal of Journalism & Communication, 41(7), 91–109.
[18] Engelke, K. M., Hase, V., & Wintterlin, F. (2019). On measuring trust and distrust in journalism: Reflection of the status quo and suggestions for the road ahead. Journal of Trust Research, 9(1), 66–86.
[19] Yao, Q., Hou, M., Fu, M., et al. (2022). The impact of bullet comments on trust in mainstream media: A user–media matching perspective. Psychological Science, 45(2), 462–469.
[20] Wang, S., Li, Z., Chen, Y., et al. (2022). Application of survey experiments in sociology: A methodological review. Sociological Review of China, 10(6), 230–252.