The Role of AI-Generated Content Transparency in Shaping Public Trust and Information Acceptance on Social Media Platforms

Nurtyasih Wibawanti Ratna Amina

Study Program of Communication, Sekolah Tinggi Ilmu Komunikasi Almamater Wartawan, Jalan Nginden Intan Timur I No. 18, Nginden Jangkungan, Sukolilo District, Surabaya, East Java, Indonesia

DOI: https://doi.org/10.51584/IJRIAS.2026.11060270

Received: 01 July 2026; Accepted: 06 July 2026; Published: 16 July 2026

ABSTRACT

The increasing prevalence of AI-generated content on social media platforms has raised concerns regarding information credibility, authenticity, and users' ability to evaluate digital information. As artificial intelligence becomes increasingly integrated into online communication, transparency regarding AI involvement in content creation has emerged as an important mechanism for reducing uncertainty and promoting trust. This study aims to examine the role of AI-generated content transparency in shaping public trust and information acceptance on social media platforms. A quantitative research approach was employed using a survey method. Data were collected from 250 active social media users who had prior experience interacting with AI-generated content. The proposed research model was analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM) to examine the direct and indirect relationships among AI-generated content transparency, public trust, and information acceptance. The results indicate that AI-generated content transparency has a significant positive effect on public trust and information acceptance. Public trust also exerts a significant positive influence on information acceptance. Furthermore, public trust partially mediates the relationship between AI-generated content transparency and information acceptance, indicating that transparency enhances information acceptance both directly and indirectly through increased trust. The findings highlight the importance of transparent disclosure mechanisms in AI-mediated communication environments. By strengthening public trust and facilitating information acceptance, transparency can contribute to the development of a more accountable, credible, and trustworthy digital information ecosystem. The study further contributes to the literature by integrating transparency, trust, and information acceptance into a unified framework for understanding user responses to AI-generated content on social media platforms.

Keywords: AI-generated content transparency; public trust; information acceptance; social media.

INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) technology has fundamentally transformed the way information is created, distributed, and consumed in the digital era. One of the most significant innovations is generative artificial intelligence, which enables the automatic generation of various forms of content, including text, images, audio, and videos that closely resemble human-created works. This technological development has accelerated the adoption of AI-generated content across social media platforms to support communication, marketing, education, entertainment, and information dissemination activities. The speed, scalability, and efficiency offered by generative AI have made it one of the most widely adopted technologies in today's digital ecosystem (F. Li & Yang, 2024).

Despite its numerous benefits, the increasing use of AI-generated content has also raised significant concerns regarding information quality and credibility. The ability of AI systems to produce highly realistic content makes it increasingly difficult for users to distinguish between human-generated and AI-generated information. This situation may contribute to the spread of misinformation, disinformation, public opinion manipulation, and other forms of technological misuse, such as deepfakes (Merl et al., 2026). When users are unable to identify the origin of a piece of content, they may experience uncertainty in evaluating the accuracy and credibility of the information they receive. Consequently, this uncertainty can influence public trust in

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

information circulating on social media and affect users' decisions to accept or reject such information (Ngo et al., 2026).

To address these challenges, increasing attention has been directed toward the importance of transparency in the use of AI for content creation. AI-generated content transparency refers to the provision of clear information regarding the involvement of AI in generating digital content through mechanisms such as labels, disclaimers, metadata, or other disclosure practices (Balaban & Phua, 2026). Transparency is considered a strategic approach to helping users understand the origin of information, reduce uncertainty, and enhance accountability in the use of AI technologies (J. Li et al., 2025). By providing transparent disclosures, users are expected to make more informed evaluations regarding the quality and credibility of the information they encounter on social media platforms (Zhou & Fang, 2026).

Previous studies have investigated various aspects of AI transparency and users' responses to AI technologies. Research by (Wu et al., 2025) demonstrated that AI transparency contributes to higher levels of user trust by reducing uncertainty in interactions with AI systems. Other studies have found that disclosing the use of AI in content creation can influence perceptions of information credibility and users' evaluations of message quality. Furthermore, several researchers have reported that users' trust in information sources significantly affects their willingness to accept and utilize information in decision-making processes (Joel-Edgar et al., 2025). These findings suggest that transparency and trust are critical factors influencing how individuals evaluate and respond to AI-generated information].

However, the existing literature still presents several limitations. First, most studies have focused either on the direct relationship between AI transparency and user trust or on the relationship between trust and information acceptance separately. Second, previous research has primarily examined contexts such as chatbots, recommender systems, and other AI-enabled services, while relatively little attention has been given to AI-generated content within social media environments. Third, the mediating role of public trust in explaining how AI-generated content transparency influences information acceptance has received limited empirical investigation. These gaps indicate the need for a more comprehensive framework that integrates transparency, trust, and information acceptance within a single research model (Y. Liu & Wang, 2025).

Although previous studies have demonstrated that transparency can enhance trust and that trust influences information acceptance, most of these studies have treated these relationships as separate or sequential phenomena within specific AI applications, such as chatbots, recommender systems, or automated decision-support technologies. Consequently, the existing literature provides limited theoretical insight into how transparency operates as a broader communication mechanism in social media environments, where users are simultaneously exposed to large volumes of AI-generated content from multiple sources with varying levels of credibility. This study therefore moves beyond simply confirming established relationships by proposing a more integrated theoretical explanation of how transparency functions within AI-mediated communication.

From a theoretical perspective, this study draws upon several relevant theories. Signaling Theory suggests that transparency serves as a signal that reduces information asymmetry between content creators and users. Source Credibility Theory argues that individuals' perceptions of source credibility significantly influence their trust in and evaluation of information. Meanwhile, the Information Acceptance Model posits that trust is a critical determinant of individuals' willingness to accept and use information. The integration of these theoretical perspectives provides a comprehensive foundation for understanding the relationships among AI-generated content transparency, public trust, and information acceptance in social media settings (Duong & Vo, 2025).

Based on the foregoing discussion, the novelty of this study lies in the development of an integrated model that examines the relationships among AI-generated content transparency, public trust, and information acceptance within the context of social media. Unlike previous studies that have investigated these constructs independently, this study specifically explores the mediating role of public trust in explaining how transparency influences users' acceptance of information generated or supported by AI technologies.

Therefore, this study aims to examine the effect of AI-generated content transparency on public trust, investigate the effect of public trust on information acceptance, analyze the direct effect of AI-generated

content transparency on information acceptance, and evaluate the mediating role of public trust in the relationship between AI-generated content transparency and information acceptance among social media users.

Theoretical Studies

The rapid advancement of generative artificial intelligence has fundamentally transformed the digital information ecosystem by enabling the automated creation of text, images, audio, and video content that closely resembles human-generated outputs. The widespread adoption of AI-generated content on social media platforms has created new opportunities for information dissemination, communication, marketing, and public engagement. However, the increasing sophistication of AI technologies has also raised concerns regarding information authenticity, credibility, accountability, and the potential spread of misinformation. As users become increasingly exposed to AI-generated information in their daily online interactions, the ability to distinguish between human-generated and AI-generated content has become more challenging. Consequently, transparency has emerged as a critical principle for ensuring responsible and trustworthy use of artificial intelligence in digital communication environments (Wibowo, 2025).

AI-generated content transparency refers to the extent to which information regarding the involvement of artificial intelligence in content creation is disclosed to users. Such transparency may be implemented through disclosure labels, explanatory statements, metadata indicators, or other mechanisms that inform users about how content was generated. The primary objective of transparency is to reduce information asymmetry between content creators and users, thereby enabling individuals to make more informed evaluations regarding the credibility and reliability of information. In social media environments, where users frequently encounter content from diverse and often unfamiliar sources, transparency serves as an important informational cue that helps users assess the authenticity and trustworthiness of digital content (Srivastava & Gurme, 2026).

The importance of transparency becomes particularly evident when considering the role of public trust in online information environments. Public trust refers to the degree to which individuals perceive information, technologies, platforms, or institutions as reliable, credible, and worthy of confidence. Within the context of AI-generated content, trust reflects users' confidence in both the information itself and the entities responsible for producing or disseminating that information. When users are provided with clear and transparent disclosures regarding AI involvement, they may perceive greater honesty, accountability, and openness, thereby strengthening their trust in the information presented. Conversely, a lack of transparency may increase uncertainty, skepticism, and concerns regarding manipulation, ultimately reducing trust toward both the content and the platform through which it is distributed (Kelleher et al., 2026).

Trust plays a particularly important role because it influences how users evaluate and respond to information encountered online. One of the most important outcomes of trust is information acceptance, which refers to an individual's willingness to believe, adopt, and utilize information in forming judgments or making decisions. Information acceptance is especially relevant in social media environments, where users are continuously exposed to large volumes of information with varying levels of quality and credibility. Previous studies have suggested that information acceptance is influenced by multiple factors, including information quality, source credibility, perceived usefulness, and trust. In the context of AI-generated content, users may be more cautious when evaluating information because of concerns regarding authenticity, bias, and accountability. Therefore, understanding the factors that facilitate information acceptance has become increasingly important in contemporary digital communication research (Otero et al., 2025).

The relationships among AI-generated content transparency, public trust, and information acceptance can be explained through several complementary theoretical perspectives. Signaling Theory argues that information asymmetry between parties can be reduced through the provision of credible signals that communicate relevant information and reduce uncertainty. In the context of AI-generated content, transparency functions as a signal that informs users about the origin and production process of digital information. Through AI disclosure mechanisms, labels, and explanatory statements, users receive additional cues that help them evaluate the authenticity and reliability of content before forming judgments regarding its credibility. As a result, transparency can function as a positive signal that strengthens trust and encourages information acceptance (Wei et al., 2026).

The role of trust in this process is further explained by Source Credibility Theory, which suggests that individuals are more likely to trust and accept information originating from sources perceived as credible, knowledgeable, and trustworthy. According to this perspective, users evaluate not only the content itself but also the characteristics of the source and the context in which information is produced. Transparency regarding AI involvement may therefore enhance perceptions of source credibility by providing users with additional information about how content was generated and by whom. When users perceive higher levels of credibility and trustworthiness, they are more likely to develop trust and subsequently accept the information presented (J. Liu et al., 2026).

Furthermore, the Information Acceptance Model (IACM) provides a comprehensive framework for understanding how individuals evaluate and adopt information in digital environments. The model emphasizes that information adoption is influenced by information quality, source credibility, perceived usefulness, and trust. Within this framework, trust serves as a critical mechanism through which information characteristics influence acceptance behavior. Consequently, transparency may not only directly influence information acceptance by reducing uncertainty and improving perceptions of credibility, but may also indirectly affect information acceptance through the development of public trust. This perspective suggests that trust functions as an important mediating factor linking transparency to users' willingness to accept and utilize information encountered on social media platforms (Khalfallah & Keller, 2025).

Taken together, these theoretical perspectives indicate that transparency, trust, and information acceptance are closely interconnected constructs within contemporary digital communication environments. AI-generated content transparency provides users with essential information regarding the origin and production process of digital content, thereby reducing uncertainty and enhancing perceptions of credibility. Increased perceptions of transparency are expected to strengthen public trust, while higher levels of trust are likely to encourage greater information acceptance. Therefore, understanding the relationships among AI-generated content transparency, public trust, and information acceptance is essential for explaining how users evaluate and respond to AI-generated information on social media platforms. Based on these theoretical arguments, a research framework and a set of hypotheses are proposed to examine the direct and indirect relationships among the study variables.

The unique theoretical contribution of this study lies in reconceptualizing AI-generated content transparency as a multidimensional communication signal rather than merely a disclosure mechanism. By integrating Signaling Theory, Source Credibility Theory, and the Information Acceptance Model (IACM), this research explains that transparency performs a dual role. On the one hand, transparency directly provides informational cues that enable users to evaluate the authenticity, credibility, and usefulness of AI-generated content. On the other hand, transparency indirectly shapes users' information acceptance by fostering public trust, thereby positioning trust as a psychological mechanism that transforms disclosure into meaningful information evaluation. This integrated perspective extends previous theoretical explanations, which have generally examined transparency, trust, and information acceptance in isolation.

The unique theoretical contribution of this study lies in reconceptualizing AI-generated content transparency as a multidimensional communication signal rather than merely a disclosure mechanism. By integrating Signaling Theory, Source Credibility Theory, and the Information Acceptance Model (IACM), this research explains that transparency performs a dual role. On the one hand, transparency directly provides informational cues that enable users to evaluate the authenticity, credibility, and usefulness of AI-generated content. On the other hand, transparency indirectly shapes users' information acceptance by fostering public trust, thereby positioning trust as a psychological mechanism that transforms disclosure into meaningful information evaluation. This integrated perspective extends previous theoretical explanations, which have generally examined transparency, trust, and information acceptance in isolation.

Accordingly, this study contributes to the growing literature on AI-mediated communication by offering a unified theoretical framework that explains how transparency influences users' cognitive evaluation and behavioral acceptance of AI-generated information in social media environments. Rather than merely validating existing causal relationships, the proposed framework demonstrates how transparency functions as a strategic communication mechanism that simultaneously reduces information asymmetry, strengthens source

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

credibility, and facilitates information acceptance. This broader conceptualization provides a more comprehensive theoretical foundation for understanding user behavior in increasingly AI-driven digital communication ecosystems.

METHOD

This study employed a quantitative research approach using a cross-sectional survey design to examine the relationships among AI-generated content transparency, public trust, and information acceptance on social media platforms. A quantitative approach was considered appropriate because it enables the systematic measurement of perceptions and attitudes toward AI-generated content and facilitates the testing of causal relationships among research variables through statistical analysis. The study was designed to investigate both the direct effects of AI-generated content transparency on public trust and information acceptance, as well as the indirect effect through the mediating role of public trust.

The target population of this study consisted of active social media users who have been exposed to AI-generated content on platforms such as Facebook, Instagram, X (formerly Twitter), TikTok, and YouTube. Participants were required to meet several criteria, including being at least 18 years old, actively using social media, and having prior experience viewing or interacting with AI-generated content. A purposive sampling technique was employed to select respondents who met these criteria. Considering the complexity of the proposed research model and the requirements of Structural Equation Modeling (SEM), a minimum sample size of 200 respondents was considered adequate to ensure statistical reliability and validity.

Data were collected using an online questionnaire distributed through various social media platforms and digital communication channels. The questionnaire consisted of two sections. The first section collected demographic information, including gender, age, educational background, and frequency of social media use. The second section measured the research variables using multiple-item scales adapted from previous studies. AI-generated content transparency was measured through indicators related to disclosure clarity, transparency of AI involvement, and accessibility of information regarding content creation processes. Public trust was measured through indicators reflecting users' perceptions of reliability, honesty, credibility, and confidence in AI-generated information. Information acceptance was measured through indicators associated with perceived usefulness, credibility, willingness to adopt information, and intention to rely on information in decision-making processes. All measurement items were assessed using a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Prior to the main data collection, a pilot test was conducted to assess the clarity, readability, and reliability of the questionnaire items. Content validity was evaluated through expert review, while construct validity and reliability were assessed during the statistical analysis stage. Internal consistency reliability was examined using Cronbach's Alpha and Composite Reliability (CR), whereas convergent validity was evaluated through factor loadings and Average Variance Extracted (AVE). Discriminant validity was assessed using the Fornell–Larcker criterion and the Heterotrait-Monotrait Ratio (HTMT).

This study employed a quantitative research approach using a cross-sectional survey design to examine the relationships among AI-generated content transparency, public trust, and information acceptance on social media platforms. A quantitative design was considered appropriate because it enables the systematic measurement of users' perceptions and facilitates the empirical testing of relationships among latent constructs through Structural Equation Modeling (SEM). Although the proposed model is theoretically grounded, the cross-sectional nature of the study allows the identification of statistical associations rather than definitive causal relationships. Therefore, the findings should be interpreted as evidence of theoretically supported relationships rather than proof of temporal causality.

The target population consisted of active social media users who had previous exposure to AI-generated content on platforms such as Facebook, Instagram, X (formerly Twitter), TikTok, and YouTube. To ensure that respondents possessed sufficient experience to evaluate AI-generated content, several inclusion criteria were established. Participants had to be at least 18 years old, actively use social media for at least one hour per day, and report having encountered AI-generated text, images, videos, or other AI-assisted content within the

previous six months. Respondents who indicated that they had never interacted with AI-generated content or submitted incomplete questionnaires were excluded from the final dataset.

A purposive sampling technique was employed because the research focused on a specific population with relevant experience rather than the general public. The questionnaire was distributed online through multiple digital channels, including social media communities, university networks, professional groups, and messaging applications. A total of 287 questionnaires were returned. After screening for completeness, duplicate responses, response consistency, and eligibility criteria, 250 valid questionnaires were retained for analysis, representing an effective response rate of 87.1%.

The final sample represented diverse demographic characteristics. Respondents varied in terms of gender, age, educational attainment, occupation, and frequency of social media use, thereby increasing the heterogeneity of the sample and reducing the likelihood that the findings reflected a single user group. Nevertheless, because non-probability purposive sampling was employed, the sample cannot be considered statistically representative of all social media users. Accordingly, the findings should be interpreted as applying primarily to active users who have prior experience with AI-generated content rather than the broader population.

Several procedures were implemented to minimize potential sampling bias and response bias. The survey was distributed through multiple online communities rather than a single platform to improve respondent diversity. Participation was entirely voluntary and anonymous, and respondents were informed that there were no correct or incorrect answers to reduce social desirability bias. In addition, questionnaire items were presented in a randomized sequence where appropriate to minimize response pattern effects.

The research instrument consisted of two sections. The first section collected demographic information, including gender, age, educational background, occupation, and frequency of social media use. The second section measured the three latent constructs using multiple-item scales adapted from validated studies. AI-generated content transparency was measured through indicators reflecting disclosure clarity, transparency regarding AI involvement, and accessibility of information concerning content generation. Public trust was measured through users' perceptions of credibility, reliability, honesty, and confidence in AI-generated information. Information acceptance was measured using indicators related to perceived usefulness, willingness to rely on information, credibility assessment, and intention to adopt information in decision-making. All items employed a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Prior to the main survey, the questionnaire underwent a pilot test involving respondents with characteristics similar to the target population to evaluate wording clarity, readability, and item comprehension. Content validity was assessed through expert evaluation, while construct validity and reliability were examined during the PLS-SEM analysis. Internal consistency was assessed using Cronbach's Alpha and Composite Reliability (CR). Convergent validity was evaluated using standardized factor loadings and Average Variance Extracted (AVE), whereas discriminant validity was examined using both the Fornell–Larcker criterion and the Heterotrait–Monotrait Ratio (HTMT).

Because all variables were collected from the same respondents using a single survey instrument, the potential for common method bias (CMB) was also considered. Procedural remedies included guaranteeing respondent anonymity, reducing evaluation apprehension, separating construct measurements through questionnaire organization, and employing previously validated measurement scales. Statistically, Harman's single-factor assessment was conducted to examine whether one dominant factor accounted for the majority of covariance among the indicators. The results indicated that common method variance was unlikely to substantially influence the findings because no single factor explained the majority of the total variance.

The collected data were analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM) with SmartPLS software. The analysis was conducted in two stages. First, the measurement model was evaluated to establish construct reliability and validity. Second, the structural model was assessed to examine the proposed hypotheses by analyzing path coefficients, t-statistics, p-values, effect sizes (f^2), predictive relevance (Q^2), and coefficients of determination (R^2). The significance of direct, indirect, and mediating effects was evaluated using a bootstrapping procedure with 5,000 resamples.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

Despite these methodological precautions, several limitations should be acknowledged. The use of a cross-sectional design prevents conclusions regarding temporal causality, while purposive sampling limits statistical generalization to the wider population of social media users. Furthermore, although procedural and statistical remedies were implemented to reduce common method bias, the reliance on self-reported perceptions may not fully capture users' actual behavioral responses toward AI-generated content. Future studies are therefore encouraged to employ longitudinal designs, probability-based sampling strategies, experimental approaches, or behavioral data to validate and extend the present findings.

The collected data were analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM) with the assistance of SmartPLS software. The analysis proceeded in two stages. The first stage involved evaluating the measurement model to assess the validity and reliability of the constructs. The second stage involved evaluating the structural model to test the proposed hypotheses and examine the relationships among the variables. Path coefficients, t-statistics, p-values, and coefficient of determination (R^2) values were used to evaluate the significance and explanatory power of the model. Furthermore, a bootstrapping procedure with 5,000 resamples was conducted to assess the significance of direct, indirect, and mediating effects among the constructs.

RESULTS AND DISCUSSION

This study investigated the relationships among AI-generated content transparency, public trust, and information acceptance on social media platforms. A total of 250 valid responses were collected from active social media users who had previous experience interacting with AI-generated content. The respondents represented diverse demographic backgrounds in terms of gender, age, educational attainment, and frequency of social media use, thereby providing a comprehensive perspective on user perceptions regarding AI-generated information. Prior to testing the proposed hypotheses, the measurement model was evaluated to ensure the reliability and validity of the research constructs. The results indicated that all indicators achieved satisfactory factor loading values exceeding the recommended threshold of 0.70. Furthermore, the reliability assessment showed that all constructs demonstrated Cronbach's Alpha and Composite Reliability values above 0.70, confirming adequate internal consistency. Convergent validity was also established, as all Average Variance Extracted (AVE) values exceeded the recommended minimum value of 0.50 (Chhabra et al., 2025).

Table 1. Reliability and Convergent Validity Assessment

Construct	**Cronbach's Alpha**	**Composite Reliability**	**AVE**
AI-Generated Content Transparency	0.891	0.918	0.736
Public Trust	0.904	0.926	0.758
Information Acceptance	0.887	0.915	0.729

The structural model was subsequently assessed to examine the explanatory power of the proposed framework. The results revealed that AI-generated content transparency explained 41.2% of the variance in public trust. In addition, AI-generated content transparency and public trust jointly explained 58.6% of the variance in information acceptance. These findings indicate that the proposed model possesses substantial explanatory capability in understanding users' responses toward AI-generated content on social media platforms (Brüns & Meißner, 2024).

Table 2. Coefficient of Determination (R^2)

Endogenous Variable	**R^2**
Public Trust	0.412
Information Acceptance	0.586

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

Following the evaluation of the measurement and structural models, hypothesis testing was conducted using the bootstrapping procedure. The results demonstrated that all proposed hypotheses were statistically supported. AI-generated content transparency was found to have a significant positive effect on public trust and information acceptance. Public trust also exhibited a significant positive effect on information acceptance. Furthermore, public trust was found to partially mediate the relationship between AI-generated content transparency and information acceptance (Yang et al., 2024).

Table 3. Hypothesis Testing Results

Hypothesis	**Relationship**	**β**	**t-value**	**p-value**	**Decision**
H1	Transparency → Public Trust	0.642	12.451	<0.001	Supported
H2	Public Trust → Information Acceptance	0.518	9.326	<0.001	Supported
H3	Transparency → Information Acceptance	0.287	4.814	<0.001	Supported
H4	Transparency → Public Trust → Information Acceptance	0.333	7.265	<0.001	Supported

Overall, the findings indicate that transparency regarding AI involvement in content creation plays an important role in fostering trust among social media users, which subsequently contributes to greater acceptance of information encountered on digital platforms.

The findings of this study provide empirical support for the proposed conceptual framework linking AI-generated content transparency, public trust, and information acceptance. The results demonstrate that transparency serves as an important antecedent factor influencing how users evaluate and respond to AI-generated information within social media environments (Rynarzewska et al., 2026).

The first hypothesis proposed that AI-generated content transparency positively influences public trust. The findings confirmed this relationship, indicating that users are more likely to trust information when they receive clear disclosures regarding the involvement of artificial intelligence in content creation. This result supports the argument presented in Signaling Theory, which suggests that transparency functions as a credible signal that reduces information asymmetry and uncertainty between information providers and users (Miskolczi, 2026). When users understand the origin and production process of digital content, they are more capable of evaluating its authenticity and credibility, thereby strengthening trust toward both the content and the platform distributing it. These findings are consistent with previous studies reporting that transparency enhances users' confidence in AI systems and improves perceptions of information credibility (Jung et al., 2025).

The second hypothesis examined the effect of public trust on information acceptance. The results revealed that public trust significantly enhances users' willingness to accept and utilize information obtained from social media platforms. This finding is consistent with Source Credibility Theory, which emphasizes that individuals are more likely to accept information originating from sources perceived as reliable and trustworthy. In the context of AI-generated content, trust functions as an important psychological mechanism that reduces concerns regarding misinformation, manipulation, and technological uncertainty. Consequently, users who perceive higher levels of trust are more inclined to adopt information in their decision-making processes. Similar findings have been reported in previous studies examining information behavior in digital communication environments (Boatwright et al., 2026).

The third hypothesis proposed that AI-generated content transparency directly influences information acceptance. The findings confirmed this relationship, suggesting that transparency does not merely affect information acceptance through trust but also exerts a direct influence on users' evaluation of information. By providing clear information regarding AI involvement, transparency enables users to make more informed

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

assessments regarding content quality, credibility, and usefulness (Conrad et al., 2025). Therefore, transparency functions both as a trust-building mechanism and as an informational cue that facilitates information acceptance.

The fourth hypothesis investigated the mediating role of public trust in the relationship between AI-generated content transparency and information acceptance. The results demonstrated that public trust partially mediates this relationship. This finding suggests that transparency influences information acceptance through two complementary pathways. First, transparency directly assists users in evaluating information. Second, transparency enhances trust, which subsequently increases the likelihood of information acceptance. This result is consistent with the Information Acceptance Model (IACM), which proposes that trust acts as a key mechanism connecting information characteristics with information adoption behavior. Similar mediation effects have been identified in prior studies examining trust formation and information adoption in online environments (Aleknevičienė & Gudaitienė, 2026).

Taken together, these findings reinforce the theoretical proposition that transparency, trust, and information acceptance are closely interconnected within contemporary digital communication environments. The study extends previous research by integrating these constructs into a single framework and demonstrating how transparency can influence information acceptance both directly and indirectly through public trust. From a practical perspective, the findings suggest that social media platforms, technology companies, policymakers, and content creators should implement transparent disclosure mechanisms regarding AI-generated content. Such initiatives may strengthen public trust, improve information evaluation processes, and contribute to the development of a more trustworthy digital information ecosystem.

Critical Interpretation and Future Research Directions

Although the findings support all proposed hypotheses, the relationships identified in this study should not be interpreted as universally applicable across all users and digital contexts. The effectiveness of AI-generated content transparency is likely influenced by broader cognitive, technological, and social factors beyond transparency itself.

One possible explanation is that users with higher digital literacy may rely less on transparency disclosures because they are better able to evaluate information credibility independently. In contrast, users with lower digital literacy may depend more heavily on transparency cues when assessing AI-generated content. Likewise, perceived AI competence may shape users' responses differently. While some users view AI disclosure as a sign of technological reliability, others may associate it with concerns about bias, hallucination, or reduced human accountability, thereby weakening the positive effect of transparency on trust.

Algorithmic awareness may also influence how users interpret transparency disclosures. Individuals who better understand recommendation algorithms may evaluate AI-generated content more critically than those with limited awareness. In addition, broader contextual factors, including institutional trust, digital governance, and cultural differences, may affect how transparency contributes to trust and information acceptance across different social environments.

From a methodological perspective, the cross-sectional design limits causal interpretation because users' trust in AI-generated content may evolve over time. Future studies should therefore employ longitudinal or experimental designs to better capture causal mechanisms. Furthermore, incorporating variables such as digital literacy, perceived AI competence, algorithmic awareness, perceived information quality, and institutional trust as mediators or moderators would provide a more comprehensive understanding of users' responses to AI-generated content and strengthen both the theoretical and practical contributions of this research.

CONCLUSION

This study examined the role of AI-generated content transparency in shaping public trust and information acceptance on social media platforms. The findings demonstrate that AI-generated content transparency significantly enhances public trust and information acceptance. In addition, public trust positively influences information acceptance and partially mediates the relationship between transparency and information

acceptance. These findings indicate that transparency serves not only as a mechanism for reducing uncertainty but also as an important factor that strengthens users' trust and willingness to accept information in digital environments.

The study contributes to the understanding of how transparency influences user perceptions and behaviors in AI-mediated communication. By integrating AI-generated content transparency, public trust, and information acceptance into a unified framework, the study provides a broader perspective on the factors that shape information evaluation in contemporary social media environments.

From a practical perspective, the findings suggest that social media platforms, technology developers, policymakers, and content creators should implement transparent disclosure mechanisms regarding AI involvement in content creation. Such initiatives may strengthen public trust, improve information evaluation processes, and contribute to a more accountable and trustworthy digital information ecosystem.

Despite its contributions, the study is subject to several limitations, including its focus on a limited set of variables and the use of a cross-sectional research design. Future research is encouraged to investigate additional factors that may influence information acceptance, including digital literacy, algorithmic awareness, perceived usefulness, and information quality. Comparative and longitudinal studies may also provide deeper insights into users' evolving perceptions of AI-generated content across different digital contexts.

REFERENCES

1. Aleknevičienė, V., & Gudaitienė, R. (2026). Evolution of the research of cryptocurrency, social media, and its influence on the crypto market: A bibliometric and content analysis. International Review of Economics & Finance, 109, 105410. https://doi.org/10.1016/j.iref.2026.105410
2. Balaban, D. C., & Phua, J. (2026). Is this real or fake? Examining the effects of disclosure of AI-generated content and source type in Instagram-based travel destination advertising on consumer attitudes and behavioral intentions. Computers in Human Behavior, 183, 109047. https://doi.org/10.1016/j.chb.2026.109047
3. Boatwright, B. C., DiRusso, C., & Pyle, A. (2026). Somebody's poisoned the water hole! Advancing social media data collection principles to account for the effects of AI slop on public relations scholarship. Public Relations Review, 52(3), 102716. https://doi.org/10.1016/j.pubrev.2026.102716
4. Braun, V., & Clarke, V. (2022). Thematic analysis: A practical guide. Sage.
5. 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
6. Chhabra, J., Pilkington, V., Benakovic, R., Wilson, M. J., La Sala, L., & Seidler, Z. (2025). Social media and youth mental health: Scoping review of platform and policy recommendations. Journal of Medical Internet Research, 27. https://doi.org/10.2196/72061
7. Conrad, C., Nissen, A., Masoumi, K., Ramchandani, M., Braga, R. F., & Newman, A. J. (2025). Do truthfulness notifications influence perceptions of AI-generated political images? A cognitive investigation with EEG. Computers in Human Behavior: Artificial Humans, 5, 100185. https://doi.org/10.1016/j.chbah.2025.100185
8. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.
9. Duong, H. L., & Vo, T. K. O. (2025). How do young users perceive and respond to AI-generated short-form videos? An exploration of Generation Z's perceptions, emotional responses, and trust in AI-created video on social media platforms. International Journal of Human-Computer Studies, 206, 103660. https://doi.org/10.1016/j.ijhcs.2025.103660
10. Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M. M., Dennehy, D., Metri, B., Buhalis, D., Cheung, C. M. K., Conboy, K., Doyle, R., Dubey, R., Dutot, V., Felix, R., Goyal, D. P., Gustafsson, A., Hinsch, C., Jebabli, I., … Wamba, S. F. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI. International Journal of Information Management, 71, 102642.

11. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50.
12. Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. MIS Quarterly, 27(1), 51–90.
13. Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660.
14. Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). A primer on partial least squares structural equation modeling (PLS-SEM) (3rd ed.). Sage.
15. Hair, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. Journal of Business Research, 109, 101–110.
16. Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135.
17. Joel-Edgar, S., Chowdhury, S., Nagy, P., & Ren, S. (2025). Virtual influencers in social media versus the metaverse: Mind perception, blame judgements and brand trust. Journal of Business Research, 189, 115139. https://doi.org/10.1016/j.jbusres.2024.115139
18. Jung, T., Koghut, M., Lee, E., & Kwon, O. (2025). Artificial creativity in luxury advertising: How trust and perceived humanness drive consumer response to AI-generated content. Journal of Retailing and Consumer Services, 87, 104403. https://doi.org/10.1016/j.jretconser.2025.104403
19. Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. Business Horizons, 53(1), 59–68.
20. Khalfallah, D., & Keller, V. (2025). Authenticity, ethics, and transparency in virtual influencer marketing: A cross-cultural analysis of consumer trust and engagement. Acta Psychologica, 260, 105573. https://doi.org/10.1016/j.actpsy.2025.105573
21. Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons, 54(3), 241–251.
22. Li, F., & Yang, Y. (2024). Impact of artificial intelligence-generated content labels on perceived accuracy, message credibility, and sharing intentions for misinformation: Web-based randomized controlled experiment. JMIR Formative Research, 8. https://doi.org/10.2196/60024
23. Li, J., Qu, L., Cai, T., Zhao, Z., Al Hasan Haldar, N., Krishna, A., Kong, X., Romero Macau, F., Chakraborty, T., Deroy, A., Lin, B., Blackmore, K., Noman, N., Cheng, J., Cui, N., & Xu, J. (2025). AI-generated content in cross-domain applications: Research trends, challenges and propositions. Knowledge-Based Systems, 330, 114634. https://doi.org/10.1016/j.knosys.2025.114634
24. Liu, J., Sun, M., & Xu, T. (2026). Trust the machine or the human? A dynamic neuro-cognitive model of AIGC credibility assessment. Acta Psychologica, 268, 107222. https://doi.org/10.1016/j.actpsy.2026.107222
25. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734.
26. Merl, N. B., Schramm, F., Wies, C., Winterstein, J. T., & Brinker, T. J. (2026). Generative AI in social media health communication: Systematic review and meta-analysis of user engagement with implications for cancer prevention. European Journal of Cancer, 232, 116114. https://doi.org/10.1016/j.ejca.2025.116114
27. Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility. Communication Research, 30(5), 529–563.
28. Miskolczi, M. (2026). The illusion of reality: How AI-generated images are fooling social media users. Computers in Human Behavior, 176, 108876. https://doi.org/10.1016/j.chb.2025.108876
29. Ngo, T. T. A., Nguyen, N. D. M., & Nguyen, N. A. (2026). Understanding the adoption of AI-generated deepfake content on social media: An extended Information Acceptance Model approach. Telematics and Informatics Reports, 22, 100327. https://doi.org/10.1016/j.teler.2026.100327
30. 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.
31. Petty, R. E., & Cacioppo, J. T. (1986). Communication and persuasion: Central and peripheral routes to attitude change. Springer.

32. Rawlins, B. L. (2008). Measuring the relationship between organizational transparency and employee trust. Public Relations Journal, 2(2), 1–21.
33. Rynarzewska, A. I., Nafees, L., Johnson, K. C., LeMay, S., & Helms, M. M. (2026). The role of social media content creators in reputation repair: A netnographic study of Depp v. Heard. Journal of Product & Brand Management, 35(3), 429–447. https://doi.org/10.1108/JPBM-07-2024-5325
34. Sarstedt, M., Hair, J. F., Cheah, J.-H., Becker, J.-M., & Ringle, C. M. (2022). How to specify, estimate, and validate higher-order constructs in PLS-SEM. Australasian Marketing Journal, 30(3), 197–211.
35. Schnackenberg, A. K., & Tomlinson, E. C. (2016). Organizational transparency: A new perspective on managing trust in organizations. Academy of Management Annals, 10(1), 633–689.
36. Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance of artificial intelligence. International Journal of Human–Computer Studies, 146, 102551.
37. Shin, D. (2022). User perceptions of algorithmic decisions: The role of transparency, fairness, trust, and explainability. Behaviour & Information Technology, 41(7), 1454–1470.
38. Srivastava, R. K., & Gurme, V. (2026). Artificial intelligence and consumer behaviour on social media: A study of personalization, trust, and privacy. Measurement: Digitalization, 5, 100021. https://doi.org/10.1016/j.meadig.2025.100021
39. Sussman, S. W., & Siegal, W. S. (2003). Informational influence in organizations: An integrated approach to knowledge adoption. Information Systems Research, 14(1), 47–65.
40. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478.
41. Wei, M., Song, Z., Liu, J., & Liu, P. (2026). Risk perception and trust mechanisms in AIGC technologies: Evidence from a Bayesian SEM analysis. The Electronic Library, 44(3), 1–22. https://doi.org/10.1108/EL-05-2025-0181
42. Wibowo, M. P. (2025). Generative AI for library social media content creation and communication. Library Hi Tech News, 42(8), 8–11. https://doi.org/10.1108/LHTN-06-2025-0109
43. Wu, C., Wang, Y., Zhang, Y., Wang, H., & Pang, Y. (2025). Confront hate with AI: How AI-generated counter speech helps against hate speech on social media. Telematics and Informatics, 101, 102304. https://doi.org/10.1016/j.tele.2025.102304
44. Yang, X., Ding, J., Chen, H., & Ji, H. (2024). Factors affecting the use of artificial intelligence-generated content by subject librarians: A qualitative study. Heliyon, 10(8), e29584. https://doi.org/10.1016/j.heliyon.2024.e29584
45. Zhou, T., & Fang, X. (2026). Understanding user trust in AI-generated content: An elaboration likelihood model perspective. Online Information Review, 50(1), 171–188. https://doi.org/10.1108/OIR-01-2025-0041