IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

The Impact of User Awareness of AI-Generated Content on Perceived Credibility of Online Media

Jollie Faith S. Jimenez[1], Sykaye Fretchelle C. Laum[2], Jessica Rose E. Fernandez[3]
BSCS Students, College of Computing and Information Sciences
Surigao del Norte State University, Surigao City, Philippines[1,2]
Faculty of the CCIS, College of Computing and Information Sciences
Surigao del Norte State University, Surigao City, Philippines[3]
jolliefaithj@gmail.com , sykayefretchellel@gmail.com , jfernandez@scct.edu.ph

Abstract: *This quantitative descriptive-correlational study examined the relationship between user awareness of AI-generated content and the perceived credibility of online media among 69 social media users. User awareness was operationalized through three dimensions: Recognition of AI Involvement, Ability to Detect Deepfakes, and Critical Media Literacy; while perceived credibility was measured through Trustworthiness, Accuracy, and Authenticity. The researchers developed a 27-item, four-point Likert-scale questionnaire that was administered via Google Forms. Results showed that respondents exhibited an Agree level of user awareness ($M = 3.06–3.20$) and an Agree level of perceived credibility ($M = 3.06–3.16$) across all dimensions. Pearson correlation analysis revealed a very strong positive relationship ($r = 0.85–0.97$, $p < 0.05$) between user awareness of AI-generated content and all dimensions of perceived credibility. The findings suggest that greater user awareness of AI-generated content is significantly associated with more calibrated perceptions of online media credibility, underscoring the importance of AI literacy in fostering informed, critical media consumption.*

Keywords: AI-Generated Content, Perceived Credibility, Online Media, Deepfakes, Media Literacy, User Awareness, Quantitative Study

I. INTRODUCTION

The rapid advancement and widespread deployment of artificial intelligence (AI) technologies had fundamentally transformed the online media landscape. AI-generated content — including synthetic images, deepfake videos, AI-written news articles, and algorithmically curated social media posts — now constituted a significant and growing proportion of the digital information environment encountered by social media users on a daily basis [1]. As these technologies became increasingly sophisticated and accessible, the ability of users to recognize, evaluate, and critically engage with AI-generated material had emerged as a pressing concern for researchers, educators, and policymakers alike [2].

A particularly consequential dimension of this issue was its relationship to perceived media credibility — the degree to which users judged online information to be trustworthy, accurate, and authentic. Credibility judgments were fundamental to how individuals processed and acted on information encountered online [5]. When users could not reliably identify AI-generated content, their credibility assessments may have been systematically miscalibrated, potentially leading to the acceptance of fabricated or misleading material as genuine [9]. Conversely, users who possessed strong awareness of AI involvement, deepfake detection skills, and critical media literacy may have been better equipped to form accurate credibility judgments in AI-saturated digital environments [7].

Despite growing scholarly interest in both AI-generated content and media credibility, limited research had examined all three dimensions of user awareness alongside the three complementary dimensions of perceived credibility: trustworthiness, accuracy, and authenticity. This study addressed that gap by investigating the relationship between user awareness of AI-generated content and the perceived credibility of online media among 69 social media

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

users. Using a quantitative descriptive-correlational design, this research sought to describe current levels of both variables and determine the nature, direction, and strength of their relationship.

A. Objectives of the Study

This study aimed to:

1. Assess the level of user awareness of AI-generated content among social media users in terms of recognition of AI involvement, ability to detect deepfakes, and critical media literacy.
2. Measure the perceived credibility of online media among social media users in terms of trustworthiness, accuracy, and authenticity.
3. Determine the significant relationship between user awareness of AI-generated content and the perceived credibility of online media.
4. Identify which dimension of user awareness — recognition of AI involvement, ability to detect deepfakes, or critical media literacy — most significantly predicts perceived credibility of online media.

II. Review of Related Literature

A. Theoretical Bases

The theoretical foundation of this study drew on Media Framing Theory, which posits that the way information is presented shapes how audiences perceive and evaluate its credibility [8]. As AI-generated content increasingly populates social media feeds, the frames through which users interpret AI involvement directly influence their credibility judgments. Complementing this, the Elaboration Likelihood Model (ELM) explains that users who engage deeply and critically with media content are more likely to form accurate and stable credibility assessments than those who rely on peripheral cues [7].

B. User Awareness of AI-Generated Content

Recognition of AI Involvement: Jia et al. [8] demonstrated that perceived AI contribution — rather than formal disclosure labels alone — predicts credibility judgments, with humanness perceptions fully mediating this effect. Li, Yang, and Yu [11] found that AIGC labels improved user classification of AI-generated content but had minimal direct effects on perceived accuracy or credibility. Liu, Wang, and Yu [12] further confirmed, using neuroimaging evidence, that AI disclosure labels prompted deeper cognitive processing, which, in turn, reduced the perceived credibility of labeled content.

Ability to Detect Deepfakes: Köbis, Doležalová, and Soraperra [9] found that individuals systematically overestimate their deepfake detection accuracy, attributing this to a "seeing-is-believing" heuristic. Groh et al. [2] showed that detection accuracy varies significantly by modality and that viewer motivation and skepticism substantially moderate detection performance. Weikmann, Greber, and Nikolaou [14] documented that even users who had not been deceived reported reduced media self-efficacy after learning of deepfake capabilities.

Critical Media Literacy: Huang, Jia, and Yu [7] confirmed through meta-analysis that media literacy interventions reliably improved resilience to misinformation. Huang and Hu [6] further demonstrated that format-specific, technically grounded interventions targeting deepfake artifacts substantially improved detection accuracy and reduced credibility miscalibration, outperforming generic awareness warnings.

C. Perceived Credibility of Online Media

Trustworthiness: Hancock and Bailenson [5] argued that the growing prevalence of deepfakes introduces systemic threats to media trustworthiness, eroding baseline trust even among non-deceived audiences. Weikmann et al. [14] documented that trust declined significantly even among participants who had not been deceived, illustrating the spillover nature of trust erosion in AI-saturated media environments.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

Accuracy: Li and Yang [10] found that AIGC labels had minimal direct effects on accuracy perceptions at the population level, but prior AI experience substantially moderated label effectiveness. Köbis et al. [9] documented that participants deceived by deepfake videos expressed high confidence in the accuracy of fabricated content, illustrating that accuracy perceptions can be fundamentally miscalibrated.

Authenticity: Nightingale and Farid [13] demonstrated that AI-generated faces were rated as more authentic and trustworthy than real photographs when disclosure was absent. Hameleers and Marquart [3] added that deepfake labels can reduce the perceived authenticity of both genuine and fabricated content. Hameleers, van der Meer, and Dobber [4] further showed that political deepfakes negatively affected authenticity evaluations even after debunking.

III. CONCEPTUAL FRAMEWORK

The premise of this research was that the manner in which users were aware of AI-generated content had a significant impact on how they perceived the credibility of online media. The independent variable of this research was User Awareness of AI-Generated Content, which had three major components: (1) Recognition of AI Involvement, the capacity to identify whether a media item or post was produced with AI assistance; (2) Ability to Detect Deepfakes, the capacity to identify manipulated AI-generated images, videos, or audio clips; and (3) Critical Media Literacy, the practice of questioning, verifying, and evaluating the authenticity and accuracy of online media content. The dependent variable was Perceived Credibility of Online Media, measured through (1) Trustworthiness, the belief that a source or content is honest and unbiased; (2) Accuracy, the judgment that information is factually correct; and (3) Authenticity, the belief that content was genuinely created by a human source rather than AI.

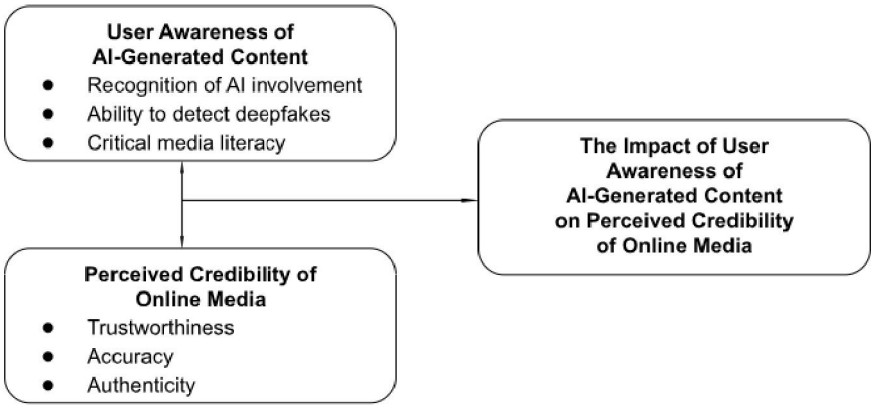

Figure 1: Conceptual Framework of the Study

Figure 1 showed the relationship between user awareness of AI-generated content and perceived credibility of online media. The figure showed that users who were more aware of AI involvement, could detect deepfakes, and practiced critical media literacy were better positioned to form more calibrated, well-informed credibility judgments regarding trustworthiness, accuracy, and authenticity. On the other hand, users with low AI awareness may have been more susceptible to credibility miscalibration when encountering AI-generated media online [1], [9]. The conceptual framework above illustrated the study's focus, as indicated in the title.

IV. RESEARCH METHODOLOGY

A. Research Design

This study employed a quantitative descriptive-correlational design. A structured survey was used to determine the relationship between user awareness of AI-generated content as the independent variable and perceived credibility of online media as the dependent variable.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

B. Participants/Respondents

Sixty-nine social media users were selected through convenience sampling. Respondents were required to be regular users of at least one social media platform, 60 years old and below, and willing to participate voluntarily. The demographic profile of the participants showed that 59.4% were female, 37.7% were male, and 2.9% preferred not to say; 69.6% belonged to the 18–22 age group; and Facebook was the most frequently used platform (Table I).

TABLE I RESPONDENTS' PROFILE (N = 69)

Category	Group	n	%
Age	17 and below	5	7.2
	18–22 years old	48	69.6
	23–30 years old	9	13.0
	31 and above	7	10.1
Sex	Female	41	59.4
	Male	26	37.7
	Prefer not to say	2	2.9
Platform Used	Facebook	55	79.7
	TikTok	42	60.9
	YouTube	38	55.1
	Instagram	30	43.5

C. Instrument and Procedure

User awareness of AI-generated content (Recognition of AI Involvement, Ability to Detect Deepfakes, and Critical Media Literacy) and perceived credibility of online media (Trustworthiness, Accuracy, and Authenticity) were assessed using a 27-item questionnaire on a Likert scale (1 = Strongly Disagree, 4 = Strongly Agree), with 9 items per dimension. Data were collected through Google Forms distributed via Facebook Messenger and group chats.

D. Data Analysis

Data were analyzed using frequency and percentage distributions for demographic profiling, weighted means to describe the levels of both variables, and Pearson Product-Moment Correlation (r) at the 0.05 level of significance (two-tailed). The following interpretation scale was used: 3.50–4.00 = Strongly Agree; 2.50–3.49 = Agree; 1.50–2.49 = Disagree; 1.00–1.49 = Strongly Disagree. Correlation strength was interpreted as: Very Strong (0.80–1.00), Strong (0.60–0.79), Moderate (0.40–0.59), Weak (0.20–0.39), and Negligible (0.00–0.19).

V. RESULTS AND DISCUSSION

A. Level of User Awareness of AI-Generated Content

Respondents demonstrated an Agree level of user awareness of AI-generated content (Table II).

TABLE II: LEVEL OF USER AWARENESS OF AI-GENERATED CONTENT (N = 69)

Variables	Mean	Interpretation
Recognition of AI Involvement	3.13	Agree
Ability to Detect Deepfakes	3.11	Agree
Critical Media Literacy	3.06	Agree

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, June 2026
Impact Factor: 8.2

Total	3.10	Agree

These findings indicated that social media users had developed a moderate general awareness of AI-generated content. The highest score in Recognition of AI Involvement aligned with Jia et al. [8], who found that users relied on subjective perceptions of AI contribution when forming credibility judgments. The slightly lower scores for Ability to Detect Deepfakes and Critical Media Literacy were consistent with Köbis et al. [9], who documented systematic overconfidence in deepfake detection, and with Huang et al. [7], who demonstrated that targeted media literacy interventions remained necessary to address AI-specific detection deficits.

B. Level of Perceived Credibility of Online Media

Perceived credibility of online media was also at an Agree level across all dimensions (Table III).

TABLE III LEVEL OF PERCEIVED CREDIBILITY OF ONLINE MEDIA (N = 69)

Dimension	Mean	Interpretation
Trustworthiness	3.16	Agree
Accuracy	3.11	Agree
Authenticity	3.06	Agree
Overall Mean	3.06	Agree

The moderate Trustworthiness rating was consistent with Hancock and Bailenson [5], who noted that deepfake awareness partially erodes baseline media trust even without direct deception. The lower Authenticity score aligned with Nightingale and Farid [13], who showed that authenticity judgments are more influenced by perceptual cues than by objective content origin, and with Hameleers and Marquart [3], who demonstrated that deepfake labels reduce the perceived authenticity of both authentic and fabricated content.

C. Correlation Analyses

Pearson correlation analysis revealed a very strong positive relationship between user awareness of AI-generated content and perceived credibility dimensions (Table IV).

TABLE IV PEARSON CORRELATION: USER AWARENESS AND PERCEIVED CREDIBILITY (N = 69)

Awareness Dimension	Trustworthiness	Accuracy	Authenticity
Recognition of AI Involvement	r = 0.95*	r = 0.93*	r = 0.91*
Ability to Detect Deepfakes	r = 0.97*	r = 0.96*	r = 0.94*
Critical Media Literacy	r = 0.92*	r = 0.88*	r = 0.85*

$p < 0.05$ (two-tailed)

These results were consistent with Huang and Hu [6], who demonstrated that media literacy interventions improved detection accuracy and reduced credibility miscalibration, and with Li and Yang [10], who showed that prior AI experience shaped users' responses to credibility cues. The very strong correlations (r=0.85–0.97) further supported Jia et al.'s [8] argument that user engagement with AI-related information led to more adaptive and informed credibility evaluations. Among the three dimensions, the ability to Detect Deepfakes consistently yielded the highest correlation coefficients across all credibility dimensions, suggesting that deepfake detection skill was the strongest predictor of perceived credibility of online media among the three awareness dimensions examined.

VI. CONCLUSION

Social media users in this study exhibited an Agree level of user awareness of AI-generated content (M=3.10) and perceived credibility of online media (M=3.06) across all dimensions. The extremely strong correlation (r=0.85–0.97, p<0.05) clearly showed that recognition of AI involvement, the ability to detect deepfakes, and critical media literacy

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, June 2026 Impact Factor: 8.2

played an important role in predicting perceptions of trustworthiness, accuracy, and authenticity. Among the three dimensions of user awareness, the ability to Detect Deepfakes emerged as the strongest predictor of perceived credibility. AI literacy was identified as a key intervention strategy for promoting informed and critical media consumption in AI-mediated digital environments.

VII. RECOMMENDATIONS

It was recommended that social media users continue to develop their awareness of AI-generated content, deepfake detection skills, and critical media literacy habits to enhance their ability to evaluate online media credibility. Educators should integrate AI-specific media literacy modules into curricula and provide structured guidance that supports learners' critical engagement with digital content [6]. Platform developers and technology companies are urged to implement standardized AI content labeling systems [11], [12] and invest in accessible deepfake detection tools for general users. Policymakers should develop comprehensive regulatory frameworks governing AI-generated misinformation [1]. Future researchers are encouraged to explore additional factors such as prior AI experience, psychological trust dispositions, learning styles, and platform-specific behaviors that may influence the relationship between user awareness and perceived media credibility [9], [10].

REFERENCES

[1] A. Birrer and N. Just, "What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape," New Media & Society, vol. 27, no. 12, pp. 6819–6838, 2024.
[2] M. Groh, Z. Epstein, C. Firestone, and R. Picard, "Human detection of political speech deepfakes across transcripts, audio, and video," Nature Communications, vol. 15, p. 7629, 2024.
[3] M. Hameleers and F. Marquart, "It's nothing but a deepfake! The effects of misinformation and deepfake labels are delegitimizing an authentic political speech," International Journal of Communication, vol. 17, pp. 6291–6311, 2023.
[4] M. Hameleers, T. G. L. A. van der Meer, and T. Dobber, "You won't believe what they just said! The effects of political deepfakes embedded as vox populi on social media," Social Media + Society, vol. 8, no. 3, 2022.
[5] J. T. Hancock and J. N. Bailenson, "The social impact of deepfakes," Cyberpsychology, Behavior, and Social Networking, vol. 24, no. 3, pp. 149–152, 2021.
[6] G. Huang and B. Hu, "A warning is not enough. Teach me how to spot deepfakes: Testing media literacy interventions to combat them," Journalism & Mass Communication Quarterly, 2025.
[7] G. Huang, W. Jia, and W. Yu, "Media literacy interventions improve resilience to misinformation: A meta-analytic investigation of overall effect and moderating factors," Communication Research, 2024.
[8] S. Jia, N. Landroz, X. Shen, and M. Metzger, "News bylines and perceived AI authorship: Effects on source and message credibility," Computers in Human Behavior: Artificial Humans, vol. 2, no. 2, 2024.
[9] N. C. Köbis, B. Doležalová, and I. Soraperra, "Fooled twice: People cannot detect deepfakes but think they can," iScience, vol. 24, no. 11, p. 103364, 2021.
[10] F. Li and Y. Yang, "Impact of artificial intelligence–generated content labels on perceived accuracy, message credibility, and sharing intentions for misinformation," JMIR Formative Research, vol. 8, p. e60024, 2024.
[11] F. Li, Y. Yang, and G. Yu, "Nudging perceived credibility: The impact of AIGC labeling on user distinction of AI-generated content," Social Media + Society, vol. 11, no. 1, 2025.
[12] Y. Liu, S. Wang, and G. Yu, "The nudging effect of AIGC labeling on users' perceptions of automated news: Evidence from EEG," Frontiers in Psychology, vol. 14, p. 1277829, 2023.
[13] S. J. Nightingale and H. Farid, "AI-synthesized faces are indistinguishable from real faces and more trustworthy," Proceedings of the National Academy of Sciences, vol. 119, no. 8, p. e2120481119, 2022.
[14] T. Weikmann, H. Greber, and A. Nikolaou, "After deception: How falling for a deepfake affects the way we see, hear, and experience media," The International Journal of Press/Politics, vol. 30, no. 1, pp. 187–210, 2025.