Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Trust and Deception in AI-Generated Advertising: The Double-Edged Role of AI Realism, Disclosure, and Consumer AI Literacy

Hyoung-me Park
Amentoventures
Joo-Eon Jeon
eric@anyang.ac.kr

Anyang University

Article

Keywords: Generative AI, AI realism, AI disclosure, AI literacy, perceived deception, consumer trust, AI-generated advertising, AI-mediated communication

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10571464/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Trust and Deception in AI-Generated Advertising: The Double-Edged Role of AI Realism, Disclosure, and Consumer AI Literacy

Abstract

Generative artificial intelligence (AI) is blurring the boundary between human-made and machine-generated communication. As AI-generated content becomes increasingly realistic, audiences must evaluate not only what a message communicates but also how and by whom it was produced. This study examines how perceived AI realism shapes trust and deception judgments in AI-generated advertising and whether these effects depend on AI disclosure and consumers' AI literacy. A between-subjects experiment was conducted with 224 adult consumers. AI realism (high vs. low) and AI disclosure (disclosed vs. undisclosed) were manipulated in a 2 x 2 experimental design, while AI literacy was treated as a measured continuous moderator. Participants viewed an AI-generated advertisement for a fictitious dietary supplement brand and evaluated advertising trust, brand trust, and perceived deception. Conditional process analyses using PROCESS Macro Model 3 were conducted with gender and age as control variables. The findings reveal a double-edged role of AI realism. Higher AI realism increased advertising trust and brand trust across experimental conditions. Disclosure strengthened the trust-enhancing effect of AI realism, while AI literacy further shaped this effect only for brand trust. For perceived deception, higher AI realism increased deception judgments most strongly when AI use was not disclosed. When disclosure was provided, conditional probing showed that this deception-enhancing effect weakened and was not significant at higher levels of AI literacy, although the three-way interaction was not statistically significant. These results show that AI realism is not merely a positive cue of content quality. Rather, its effects

depend on transparency practices and consumers' interpretive competence regarding AI-generated content.

Keywords: Generative AI; AI realism; AI disclosure; AI literacy; perceived deception; consumer trust; AI-generated advertising; AI-mediated communication

1. Introduction

Generative artificial intelligence (AI) is rapidly changing the production and evaluation of persuasive communication. While AI has long been used to support marketing decisions such as targeting, recommendation, segmentation, and analytics, recent advances in text, image, voice, and video generation have made AI a direct producer of communication content. As AI-generated outputs become increasingly realistic, audiences may find it difficult to distinguish between content created by humans and content created by machines. This development changes the basis on which people evaluate messages. Consumers are likely to consider not only the informational content of an advertisement but also whether the message was produced by a human creator, generated by AI, or produced through human-AI collaboration. Recent advertising research has therefore begun to examine AI-created advertisements, deepfakes, automated brand-generated content, and disclosure of AI involvement as emerging issues in persuasive communication (Campbell et al., 2022; Huh et al., 2023; Van Noort et al., 2020; Wu and Wen, 2021).

This shift raises an important question for the social scientific study of AI-mediated communication: how do people form judgments of trust and deception when the boundary between human-made and machine-generated content is blurred? In traditional advertising research, consumer evaluation has been explained by factors such as informativeness, creativity, emotional appeal, source credibility, and brand fit. In AI-generated advertising, however, the

perceived production process and the realism of the resulting message may serve as additional cues. When an AI-generated advertisement looks natural, human-like, and realistic, consumers may perceive it as professional and credible. At the same time, they may also become concerned about manipulation, inauthenticity, or hidden persuasion when the role of AI is not transparent.

This study focuses on perceived AI realism, defined as the extent to which consumers perceive AI-generated advertising content as natural, human-like, realistic, and authentic. Prior work on anthropomorphism, social presence, agency, and humanness suggests that people often respond socially to nonhuman agents when they display human-like characteristics (Epley et al., 2007; Go and Sundar, 2019; Lankton et al., 2015; Reeves and Nass, 1996). Applied to AI-generated advertising, such research implies that realistic expressions, voice, movement, visual composition, and message flow may lead consumers to evaluate AI-generated content more favorably. Yet the same realism may also create discomfort or suspicion if consumers believe that a machine-generated message is imitating human communication without sufficient transparency.

We further examine two conditions that may shape how consumers interpret AI realism: AI disclosure and AI literacy. AI disclosure refers to whether consumers are explicitly informed that AI was used in creating the advertisement. Disclosure can operate as a signal of transparency and responsibility, but it can also activate persuasion knowledge and make consumers more attentive to persuasive intent or possible manipulation (Friestad and Wright, 1994; Spence, 1973). AI literacy refers to consumers' ability to understand how AI works, recognize its limitations, and critically evaluate AI-generated outputs (Long and Magerko, 2020). Research on consumer responses to AI further shows that trust, aversion, and acceptance depend on the perceived fit between AI and the task, prior experience with algorithmic error, and broader experiential

appraisals of AI systems (Castelo et al., 2019; Dietvorst et al., 2015; Glikson and Woolley, 2020; Puntoni et al., 2021). Consumers with higher AI literacy may interpret realistic AI-generated advertising as evidence of technological competence and communication quality, whereas consumers with lower AI literacy may interpret AI involvement as a source of uncertainty or deception.

This study makes three contributions. First, it identifies perceived AI realism as a central cue in AI-mediated persuasive communication, extending prior work on anthropomorphism and social presence from AI agents to AI-generated content. Second, it shows that AI realism has a double-edged role: it can enhance advertising and brand trust while also increasing perceived deception under conditions of limited transparency or lower AI literacy. Third, it demonstrates that the effects of AI-generated content depend jointly on message characteristics, disclosure practices, and consumers' interpretive competence, highlighting the importance of conditional models in research on generative AI communication.

2. Theoretical background and hypotheses

2.1 Perceived AI realism

Perceived AI realism refers to the degree to which consumers experience an AI-generated output as natural, human-like, realistic, and authentic. It goes beyond the simple recognition that a message may have been produced by AI. It includes perceptions of intentionality, emotionality, autonomy, and natural expression. As generative AI content becomes more visually and verbally sophisticated, consumers may increasingly rely on the perceived realism of a message as a cue for evaluating its credibility and persuasive quality.

AI realism is conceptually related to humanness, anthropomorphism, agency, and social presence. Humanness captures the extent to which a technological agent or output is perceived as similar to a human in appearance, language, behavior, or expression. Anthropomorphism refers to the tendency to attribute human-like qualities, intentions, or emotions to nonhuman entities (Epley et al., 2007). Agency refers to the perception that AI can act, decide, or communicate as an autonomous agent. Social presence refers to the extent to which a communication partner or mediated object is experienced as socially present (Biocca, 1997; Short et al., 1976; Tu, 2002). These concepts suggest that when AI-generated advertising provides natural language, realistic imagery, and human-like expressive cues, consumers may evaluate it as more socially meaningful and credible.

Perceived AI realism may therefore increase advertising trust and brand trust. When an AI-generated advertisement appears natural and professionally produced, consumers may infer that the advertisement provides reliable information and that the brand is capable of delivering high-quality communication. Advertising trust captures the belief that an advertisement is believable, useful, and informative, whereas brand trust reflects the belief that a brand is reliable and considerate of consumers' interests (Delgado-Ballester, 2004; MacKenzie and Lutz, 1989; Soh et al., 2009). Realistic AI-generated advertising may strengthen these forms of trust by increasing perceptions of message quality and brand competence.

At the same time, AI realism may also affect perceived deception. Perceived deception refers to the extent to which consumers believe that an advertisement is misleading, dishonest, or intended to manipulate them (Darke and Ritchie, 2007). Although realism can make an advertisement appear more professional, it can also raise concerns when the production process is unclear. If an AI-generated advertisement appears almost human-made but consumers are not informed about

AI use, the realism of the content may be interpreted as concealment or imitation. Thus, the effect of AI realism on perceived deception is likely to be conditional rather than uniformly positive or negative.

H1a. Advertisements with high AI realism will generate higher advertising trust than advertisements with low AI realism.

H1b. Advertisements with high AI realism will generate higher brand trust than advertisements with low AI realism.

H1c. AI realism will influence perceived deception.

2.2 AI disclosure

AI disclosure refers to whether consumers are explicitly informed that AI was used in the production of an advertisement. As generative AI becomes more deeply embedded in persuasive communication, the disclosure of AI use is becoming a central issue of transparency. Disclosure provides information about how a message was produced and may therefore influence how consumers judge authenticity, trust, persuasive intent, and possible deception.

From a signaling perspective, disclosure can act as a positive signal. In contexts of information asymmetry, voluntary disclosure may communicate transparency, accountability, and confidence in the quality of a message (Spence, 1973). If a company openly indicates that AI was used in creating a realistic advertisement, consumers may interpret the disclosure as evidence that the company is transparent and technologically competent. Under such conditions, AI disclosure may strengthen the positive effects of AI realism on advertising trust and brand trust.

However, disclosure can also activate persuasion knowledge. When consumers learn that a message was generated by AI, they may become more aware of the persuasive intent of the

advertisement and more attentive to the possibility that the message was strategically designed to influence them (Campbell and Kirmani, 2000; Friestad and Wright, 1994). Prior research on sponsored content and covert advertising similarly shows that disclosure can alter consumers' interpretation of persuasive messages (Boerman et al., 2017; Evans and Park, 2015). Emerging studies on AI-generated advertising further suggest that AI disclosure can affect advertising attitudes, perceived credibility, and inferences of manipulative intent, although its effects are not uniformly positive or negative (Arango et al., 2023; Wortel et al., 2024). In AI-generated advertising, disclosure may therefore operate as a mixed cue: it can enhance transparency but also make consumers more critical.

The effect of disclosure is expected to depend on the level of AI realism. When AI realism is high, disclosure may help consumers interpret realism as the outcome of sophisticated and transparent technology use. When AI realism is low, disclosure may instead draw attention to the limitations of the technology or the artificiality of the advertisement. Conversely, when AI use is not disclosed, consumers lack explicit information about the production process. In such cases, the effects of AI realism may be weaker for trust but stronger for suspicion when consumers infer that the realistic appearance of the message conceals machine production.

H2. AI disclosure will moderate the effect of AI realism on trust and deception judgments. Specifically, when AI use is disclosed, the effects of AI realism on advertising trust and brand trust will be strengthened, while its effect on perceived deception will vary across conditions.

2.3 AI literacy

AI literacy refers to the ability to understand the basic logic, capabilities, limitations, and social implications of AI, and to critically evaluate AI-generated information (Long and Magerko, 2020). In the context of AI-generated advertising, AI literacy is not merely familiarity with AI

tools. It is an interpretive competence that helps consumers evaluate how AI may have contributed to a message and whether the resulting content should be trusted.

Consumers with higher AI literacy are likely to evaluate AI-generated advertisements more analytically. They may understand that AI-generated outputs can vary in quality, that AI systems can produce errors or biases, and that the use of AI does not necessarily imply deception. When exposed to highly realistic AI-generated advertising, such consumers may consider technological quality, message coherence, product relevance, and disclosure practices. As a result, they may interpret AI realism as a positive cue of communication quality when the production process is transparent.

By contrast, consumers with lower AI literacy may have greater uncertainty about how AI-generated content is produced and what its limitations are. They may interpret AI involvement as a sign of nonhuman communication, manipulation, or reduced authenticity. Under these circumstances, highly realistic AI-generated advertising may be less consistently associated with trust and may instead trigger suspicion, particularly when the role of AI is not clearly understood.

H3. AI literacy will moderate the effect of AI realism on trust and deception judgments. Specifically, consumers with higher AI literacy will show stronger trust-related responses to high AI realism, while the effect of AI realism on perceived deception will vary across disclosure conditions.

2.4 Joint effects of AI realism, disclosure, and literacy

The evaluation of AI-generated advertising is unlikely to depend on AI realism, disclosure, or AI literacy in isolation. Consumers evaluate the visible realism of the advertisement, the information

provided about how it was produced, and their own capacity to interpret AI-generated content. These factors may jointly determine whether realistic AI-generated advertising is interpreted as professional and trustworthy or as artificial and potentially deceptive.

When AI use is disclosed and consumers have higher AI literacy, high AI realism may be interpreted as a sign of technological competence and transparent communication. In this case, AI realism and disclosure may complement one another, leading to higher advertising trust and brand trust and lower perceived deception. By contrast, when consumers have lower AI literacy, disclosure may not fully resolve uncertainty about AI-generated content. Even with disclosure, these consumers may interpret highly realistic AI advertising as human imitation or possible manipulation.

Because brand trust reflects broader judgments about organizational competence and reliability, AI literacy may be especially consequential for brand-level evaluations. Consumers with higher AI literacy may be more capable of translating disclosed AI use and high AI realism into inferences about the brand's technological competence and transparency. This suggests that the joint effect of AI realism, disclosure, and AI literacy may be particularly visible in judgments that extend beyond the advertisement itself to the sponsoring brand.

When AI use is not disclosed, the meaning of realism may also change. High realism may enhance surface-level trust by making a message appear polished and professional, but it may simultaneously increase perceived deception if consumers suspect that the message conceals its machine-generated nature. This possibility is consistent with research showing that high human likeness can sometimes produce negative reactions when it violates expectations or creates discomfort (Mathur and Reichling, 2016). Accordingly, the effect of AI realism may depend on the joint conditions of disclosure and AI literacy.

H4a. AI realism, AI disclosure, and AI literacy will have a three-way interaction effect on advertising trust.

H4b. AI realism, AI disclosure, and AI literacy will have a three-way interaction effect on brand trust.

H4c. AI realism, AI disclosure, and AI literacy will have a three-way interaction effect on perceived deception.

3. Methods

3.1 Research design and stimuli

The study used a 2 x 2 between-subjects experimental design in which AI realism (high vs. low) and AI disclosure (disclosed vs. undisclosed) were manipulated. AI literacy was measured as a continuous individual-difference moderator. The dependent variables were advertising trust, brand trust, and perceived deception.

The experimental stimulus was an advertisement for a fictitious dietary supplement brand named IMMUNEX. This product category was selected because evaluations of dietary supplements often depend on perceived efficacy, safety, ingredients, expert recommendation, and trustworthiness. The advertisement was an approximately 18-second video in which a professional pharmacist introduced the benefits of IMMUNEX and recommended the brand.

The four stimulus videos were designed to keep the advertised brand, product information, message content, length, and recommendation structure constant across conditions. The realism manipulation varied only the naturalness of the AI-generated model's facial expression, gaze, movement, voice, and scene composition. AI disclosure was manipulated by presenting, before

the advertisement, a message stating that the advertisement had been produced using artificial intelligence technology. Participants in the no-disclosure condition received no such information. The manipulation was designed to alter perceived realism rather than the substantive message, although related perceptions such as production quality and attractiveness were not independently manipulated.

A pretest with 62 participants was conducted to verify the AI realism manipulation. Participants perceived the high-realism advertisement ($M = 3.45$, $SD = 1.47$) as significantly more realistic than the low-realism advertisement ($M = 2.24$, $SD = 1.17$), $t = 4.421$, $p < 0.001$. These results indicated that the manipulation of AI realism was successful.

3.2 Participants and procedure

The main experiment was conducted through a professional online research panel. The target population consisted of adults aged between 30 and 69 years, a group relevant to the dietary supplement category and likely to have varying degrees of exposure to digital technologies and AI. A total of 224 valid responses were used for the final analysis after excluding incomplete or careless responses.

Participants were randomly assigned to one of four experimental conditions. After reading information about the study, anonymity, voluntary participation, data use, and the right to withdraw, participants provided informed consent and completed demographic questions. They then responded to items measuring AI literacy. After that, they viewed the advertisement corresponding to their assigned experimental condition. Participants in the AI disclosure condition saw a disclosure message before viewing the advertisement, whereas those in the no-disclosure condition did not. After viewing the advertisement, participants completed

manipulation-check items and measures of advertising trust, brand trust, and perceived deception.

The final sample consisted of 114 men (50.9%) and 110 women (49.1%). The age distribution was relatively balanced across adulthood: 54 participants were in their 30s (24.1%), 50 were in their 40s (22.3%), 53 were in their 50s (23.7%), and 67 were in their 60s (29.9%). In terms of education, 144 participants (64.3%) had attended or graduated from university, 48 (21.4%) had completed high school or less, and 32 (14.3%) had attended or graduated from graduate school.

3.3 Measures and manipulation checks

All constructs were measured using multi-item scales. AI realism was measured with four items assessing whether the advertisement model felt human-like, communicative, warm, and vivid. AI disclosure perception was measured with three items assessing the extent to which participants believed the advertisement was created by AI. AI literacy was measured with five items assessing understanding of how AI works, knowledge of generative AI limitations, ability to distinguish AI-generated from human-made content, ability to critically evaluate AI-generated information, and general perceived knowledge of AI. Advertising trust was measured with three items assessing the extent to which the advertisement was believable, useful, and understandable. Brand trust was measured with four items assessing whether the brand was trustworthy, reliable, promise-keeping, and consumer-oriented. Perceived deception was measured with four items assessing whether the advertisement was deceptive, misleading, false, or intended to deceive.

Internal consistency was satisfactory for all measures. Cronbach's alpha values were .967 for AI realism, 0.965 for AI disclosure perception, 0.936 for AI literacy, 0.812 for advertising trust, 0.817 for brand trust, and .813 for perceived deception. Manipulation checks also confirmed successful experimental manipulations. Participants in the AI disclosure condition (M = 5.88, SD

= .42) reported significantly greater perception of AI use than those in the no-disclosure condition (M = 1.95, SD = .37), $F(1, 222) = 5487.99$, $p < 0.001$. Participants in the high AI realism condition (M = 5.33, SD = .54) perceived the advertisement as significantly more realistic than participants in the low AI realism condition (M = 2.67, SD = .62), $F(1, 222) = 1181.19$, $p < 0.001$.

3.4 Analytic strategy

The hypotheses were tested using Hayes's (2022) PROCESS Macro Model 3. Separate conditional process analyses were conducted for advertising trust, brand trust, and perceived deception. AI realism was entered as the independent variable, AI disclosure and AI literacy were entered as moderators, and gender and age were entered as control variables. AI realism was coded so that the high-realism condition was 0 and the low-realism condition was 1. AI disclosure was coded as no disclosure = 0 and disclosure = 1. AI literacy was treated as a measured continuous moderator and mean-centered before computing interaction terms (M = 4.14, SD = 1.10). For interpretation and visualization, conditional effects were probed at lower (-1 SD) and higher (+1 SD) levels of AI literacy. Because models with interaction terms make main effects conditional on reference values of the moderators, interpretation focuses primarily on two-way interactions, three-way interactions, and conditional effects.

4. Results

4.1 Conditional process models

Table 1 presents the conditional process models for the three dependent variables. For advertising trust, the overall model was statistically significant and explained 63.8% of the

variance, $F = 41.995$, $p < 0.001$, $R^2 = 0.638$, adjusted $R^2 = 0.623$. AI disclosure had a positive effect on advertising trust, whereas gender and age were not significant. The effect of AI realism was significant when probed conditionally: lower AI realism produced lower advertising trust than higher AI realism. The interaction between AI realism and AI disclosure was significant, $B = -0.374$, $SE = 0.103$, $t = -3.634$, $p < 0.001$, indicating that disclosure strengthened the trust-enhancing effect of AI realism. The three-way interaction with AI literacy was not statistically significant, $B = -0.154$, $SE = 0.094$, $t = -1.638$, $p = 0.103$. Thus, H4a was not supported in the continuous-moderator model.

For brand trust, the overall model was statistically significant and explained 79.7% of the variance, $F = 93.435$, $p < 0.001$, $R^2 = 0.797$, adjusted $R^2 = 0.789$. AI realism, AI disclosure, and AI literacy were significant predictors of brand trust. The interactions between AI realism and disclosure, between AI disclosure and AI literacy, and the three-way interaction among AI realism, disclosure, and literacy were significant. The three-way interaction was $B = -0.341$, $SE = 0.100$, $t = -3.397$, $p = 0.001$. These results indicate that the effect of AI realism on brand trust depended jointly on disclosure and AI literacy. Thus, H4b was supported.

For perceived deception, the overall model was statistically significant and explained 79.3% of the variance, $F = 90.859$, $p < 0.001$, $R^2 = 0.793$, adjusted $R^2 = 0.784$. AI realism and AI disclosure had significant effects on perceived deception, whereas AI literacy did not. The interaction between AI realism and disclosure was significant, $B = 1.105$, $SE = 0.102$, $t = 10.853$, $p < 0.001$, indicating that the deception-enhancing effect of AI realism varied substantially by disclosure condition. However, the three-way interaction among AI realism, disclosure, and literacy was not statistically significant, $B = 0.147$, $SE = 0.093$, $t = 1.574$, $p = 0.117$. Thus, H4c

was not supported, although conditional probing suggested that the deception-enhancing effect of AI realism was attenuated at higher levels of AI literacy when disclosure was provided.

Table 1. Conditional process models predicting trust and deception judgments

Variable	Advertising trust B (SE)	Brand trust B (SE)	Perceived deception B (SE)
Constant	4.490*** (0.143)	3.676*** (0.153)	5.688*** (0.142)
Gender	-0.047 (0.052)	0.126* (0.055)	-0.013 (0.051)
Age	-0.003 (0.002)	-0.000 (0.002)	-0.004 (0.002)
AI realism (low vs. high)	-0.651*** (0.072)	-0.590*** (0.077)	-1.401*** (0.072)
AI disclosure	0.658*** (0.074)	1.423*** (0.079)	-1.526*** (0.073)
AI literacy (mean-centered)	-0.038 (0.046)	0.069 (0.049)	-0.039 (0.046)
AI realism x AI disclosure	-0.374*** (0.103)	-1.022*** (0.110)	1.105*** (0.102)
AI realism x AI literacy	0.103 (0.066)	-0.000 (0.070)	0.060 (0.065)
AI disclosure x AI literacy	0.166* (0.065)	0.226** (0.070)	-0.166* (0.065)
AI realism x AI disclosure x AI literacy	-0.154 (0.094)	-0.341*** (0.100)	0.147 (0.093)
F	41.995	93.435	90.859
R^2	0.638	0.797	0.793
Adjusted R^2	0.623	0.789	0.784

*Note. Standard errors are in parentheses. AI realism was coded high = 0 and low = 1; AI disclosure was coded no disclosure = 0 and disclosure = 1; AI literacy was mean-centered and treated as a continuous moderator. * p < .05, ** p < .01, *** p < 0.001.*

4.2 Conditional effects of AI realism

Table 2 reports the conditional effects of AI realism at lower (-1 SD) and higher (+1 SD) levels of AI literacy. Because AI realism was coded high = 0 and low = 1, negative conditional effects indicate that the low-realism condition produced lower outcome scores than the high-realism condition. For advertising trust and brand trust, this means that higher AI realism increased trust. For perceived deception, this means that higher AI realism increased perceived deception.

Table 2. Conditional effects of AI realism at values of AI disclosure and AI literacy

Outcome	AI disclosure	AI literacy	Effect	SE	t	p	LLCI	ULCI
Advertising trust	No disclosure	Lower (-1 SD)	-0.764	0.103	-7.397	< 0.001	-0.967	-0.560
Advertising trust	No disclosure	Higher (+1 SD)	-0.538	0.101	-5.296	< 0.001	-0.738	-0.337
Advertising trust	Disclosure	Lower (-1 SD)	-0.969	0.105	-9.254	< 0.001	-1.175	-0.762
Advertising	Disclosure	Higher (+1	-1.081	0.101	-10.660	< 0.001	-1.280	-0.881

trust		SD)						
Brand trust	No disclosure	Lower (-1 SD)	-0.589	0.110	-5.350	< 0.001	-0.806	-0.372
Brand trust	No disclosure	Higher (+1 SD)	-0.590	0.108	-5.452	< 0.001	-0.804	-0.377
Brand trust	Disclosure	Lower (-1 SD)	-1.237	0.112	-11.079	< 0.001	-1.457	-1.017
Brand trust	Disclosure	Higher (+1 SD)	-1.986	0.108	-18.368	< 0.001	-2.199	-1.773
Perceived deception	No disclosure	Lower (-1 SD)	-1.466	0.102	-14.344	< 0.001	-1.667	-1.264
Perceived deception	No disclosure	Higher (+1 SD)	-1.335	0.100	-13.294	< 0.001	-1.533	-1.137
Perceived deception	Disclosure	Lower (-1 SD)	-0.522	0.104	-5.038	< 0.001	-0.726	-0.318
Perceived deception	Disclosure	Higher (+1 SD)	-0.069	0.100	-0.692	0.489	-0.267	0.128

Note. AI literacy values correspond to one standard deviation below and above the mean. Negative effects indicate lower scores in the low-realism condition relative to the high-realism condition.

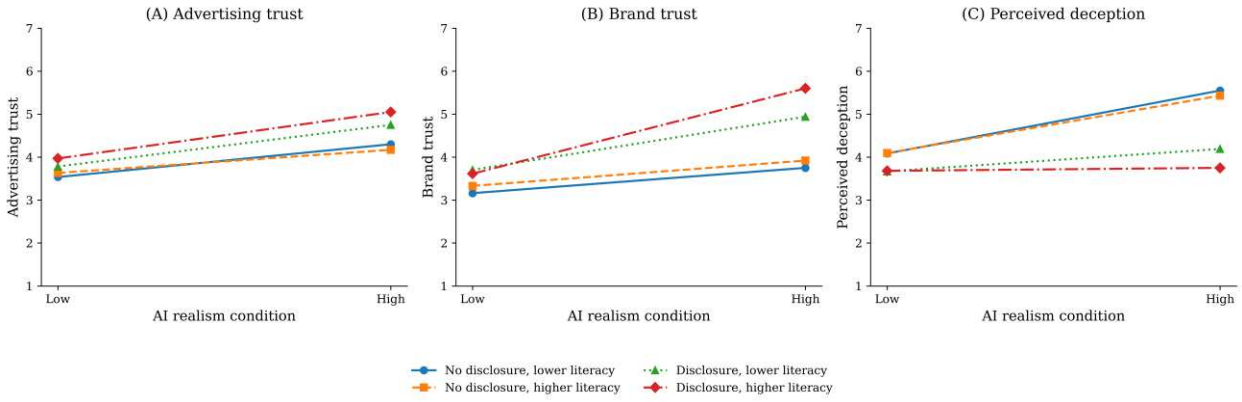
For advertising trust, AI realism had significant conditional effects across all combinations of disclosure and AI literacy. Without disclosure, the low-realism condition generated lower advertising trust than the high-realism condition at both lower and higher levels of AI literacy. With disclosure, the difference between low and high AI realism was larger, indicating that disclosure strengthened the positive effect of AI realism on advertising trust. However, because the three-way interaction was not significant, differences across levels of AI literacy should be interpreted descriptively rather than as evidence of a statistically supported moderated moderation effect.

For brand trust, AI realism also had significant conditional effects across all combinations of disclosure and AI literacy. The effect was relatively stable when AI use was not disclosed, but it became substantially stronger when AI use was disclosed. Under disclosure, the effect of AI realism increased as AI literacy increased. This pattern is consistent with the significant three-way interaction and indicates that consumers with higher AI literacy interpreted realistic AI-generated advertising more favorably when the use of AI was transparent.

For perceived deception, AI realism increased deception judgments most strongly when AI use was not disclosed. With disclosure, the effect was weaker. It remained significant at lower AI

literacy but was not significant at higher AI literacy. Thus, transparent disclosure appeared to attenuate the deception-enhancing effect of AI realism, particularly among consumers with greater interpretive competence regarding AI-generated content. Because the three-way interaction was not statistically significant, this attenuation should be interpreted as a conditional probing pattern rather than as confirmatory evidence for H4c.

Figure 1. Three-way interaction effects of AI realism, AI disclosure, and AI literacy. Note. Values are model-predicted scores from PROCESS Model 3 with AI literacy probed at ±1 SD and gender and age held at their sample means. Higher values indicate higher scores on each outcome. Lower and higher AI literacy correspond to one standard deviation below and above the mean, respectively.



[Figure 1 about here]

5. Discussion

5.1 Summary of findings

This study examined how perceived AI realism influences trust and deception judgments in AI-generated advertising and how these effects are conditioned by AI disclosure and AI literacy.

Overall, the results provide partial support for the proposed conditional framework and strong support for the double-edged role of AI realism. For advertising trust and brand trust, high AI realism consistently produced more favorable evaluations than low AI realism. For advertising trust, disclosure strengthened the effect of AI realism, although the continuous AI literacy model did not support a significant three-way interaction. For brand trust, the three-way interaction was significant: the trust-enhancing effect of AI realism was strongest when AI use was disclosed and consumers had higher AI literacy. One possible explanation for this outcome-specific pattern is that brand trust requires consumers to move beyond the immediate advertisement and infer broader organizational competence and reliability. Such inferences may require greater interpretive integration of AI realism, disclosure, and AI literacy than immediate advertising trust or perceived deception judgments.

The results for perceived deception were more complex. High AI realism increased perceived deception, particularly when AI use was not disclosed. This suggests that highly realistic AI-generated advertising can create suspicion when the production process is opaque. When AI use was disclosed, the deception-enhancing effect of AI realism became weaker and disappeared at higher levels of AI literacy. However, because the three-way interaction was not significant in the continuous-moderator model, this pattern should be interpreted cautiously as suggestive evidence that AI disclosure may work more effectively among consumers who can better interpret AI-generated content.

Taken together, the findings suggest that AI realism is not uniformly beneficial. It can enhance trust by signaling professional communication quality and technological competence, but it can also increase perceived deception when consumers perceive realistic AI-generated content as hidden imitation or insufficiently transparent machine production.

5.2 Theoretical implications

First, this study contributes to research on AI-mediated communication by identifying perceived AI realism as a key evaluative cue. Much prior research has examined whether consumers accept or resist AI, trust algorithms, or respond to AI agents. This study shifts attention to the perceptual qualities of AI-generated content itself. The findings show that the realism of AI-generated advertising is not simply a technical feature; it is a social and interpretive cue that shapes trust and deception judgments.

Second, the study advances the literature by demonstrating the double-edged nature of AI realism. Human-like and realistic technological cues have often been associated with social presence, anthropomorphism, and trust. The present findings show that such cues can also trigger suspicion when the production process is opaque. AI realism therefore should not be conceptualized only as a positive feature of message quality. Rather, it can enhance trust and, under limited transparency, increase perceived deception.

Third, the study clarifies the role of AI disclosure. Disclosure did not simply produce a uniform increase or decrease in evaluation. Instead, disclosure changed how consumers interpreted the realism of AI-generated advertising. In the trust models, disclosure strengthened the positive effects of AI realism. In the deception model, disclosure substantially reduced the deception-enhancing effect of AI realism. These results suggest that transparency is a central condition under which realistic AI-generated content is interpreted as credible rather than suspicious.

Fourth, the study positions AI literacy as an interpretive competence in AI-mediated communication. AI literacy did not uniformly moderate all outcomes, but it played a meaningful role in brand trust and in the conditional pattern of perceived deception under disclosure. This suggests that AI literacy may be especially important when consumers connect AI-generated

communication to broader judgments of organizational competence, transparency, and brand reliability.

Finally, by testing conditional models across trust and deception outcomes, the study demonstrates that consumer responses to AI-generated content are unlikely to be explained by single-factor models. Instead, responses emerge from the interaction between content characteristics, transparency practices, and individual interpretive capacity. This conditional perspective is particularly important as generative AI makes machine-generated communication increasingly difficult to distinguish from human-made communication.

5.3 Practical implications

The findings have several practical implications for organizations using generative AI in persuasive communication. First, companies should pay attention to the perceived realism of AI-generated advertising. Realistic expression, natural movement, coherent voice, and high-quality visual composition can increase advertising trust and brand trust. This may be particularly important for product categories, such as dietary supplements, in which credibility, safety, and expertise are central to consumer evaluation.

Second, organizations should not treat AI disclosure as a mere legal or technical label. The findings suggest that disclosure can strengthen the positive effects of AI realism on trust and reduce the deception-enhancing effect of realism. Rather than hiding AI use, brands may benefit from clearly explaining how AI contributed to the advertisement. For example, disclosure can describe whether AI was used for visual generation, voice synthesis, creative support, or message production. Such information may help consumers interpret AI use as transparent and responsible rather than deceptive.

Third, communication strategies should account for variation in AI literacy. Consumers with lower AI literacy may still interpret highly realistic AI advertising as deceptive even when AI use is disclosed. For these consumers, a simple statement that an advertisement was produced using AI may be insufficient. Brands may need to provide more accessible explanations of why AI was used, what role it played, and how the accuracy or quality of the message was ensured.

Fourth, the pursuit of realism should be balanced with transparency and ethical communication. The findings caution against the assumption that making AI-generated advertising indistinguishable from human-made advertising is always beneficial. Highly realistic AI content can increase trust, but it can also trigger perceived deception when consumers feel that the production process is hidden or difficult to understand. Companies should therefore design AI-generated advertising with attention to realism, disclosure, ethical clarity, and consumer comprehension.

5.4 Limitations and future research

This study has several limitations that suggest directions for future research. First, the experiment used an advertisement for a fictitious dietary supplement brand. This product category was appropriate because trust, safety, and expertise are important in evaluations of health-related products. However, the findings may not generalize to all product categories. Future research should examine whether similar patterns emerge in hedonic products, luxury goods, financial services, medical services, public communication, entertainment, and low-involvement consumer products.

Second, although the manipulation check confirmed that participants perceived the high-realism advertisement as more realistic than the low-realism advertisement, future research should further separate AI realism from related perceptions such as production quality, attractiveness,

informativeness, and message professionalism. This would help determine whether consumers are responding specifically to AI realism or to broader evaluations of production quality.

Third, AI literacy was measured rather than experimentally manipulated. Although this approach captures individual differences in perceived AI understanding, it does not establish the causal effect of AI literacy. Future research could manipulate AI literacy through educational interventions, AI training, or explanations of the AI production process to test whether improving AI literacy changes consumer responses to realistic AI-generated content.

Fourth, the study used a binary disclosure manipulation. In practice, AI disclosure may differ in wording, specificity, timing, location, and visual prominence. Future studies should compare different disclosure formats, such as 'created by AI', 'created with the assistance of AI', or 'created through human-AI collaboration'. Such research could identify disclosure strategies that maximize transparency while minimizing unnecessary suspicion.

Fifth, the study focused on advertising trust, brand trust, and perceived deception. Future research could extend the model to additional outcomes such as attitude toward the advertisement, brand attitude, purchase intention, information search, sharing intention, ad avoidance, perceived authenticity, perceived creativity, and perceived manipulative intent. Such outcomes would provide a more comprehensive understanding of how AI realism influences persuasion.

Finally, this study measured consumer responses after a single exposure to an AI-generated advertisement. In real communication environments, consumers may repeatedly encounter AI-generated content and may become more familiar with it over time. Longitudinal and repeated-exposure research could examine whether the trust-enhancing or deception-triggering effects of AI realism change as consumers gain more experience with generative AI. Future research

should also examine the psychological mechanisms underlying the double-edged effect, such as perceived authenticity, perceived manipulation, transparency perception, technological anxiety, and persuasion knowledge activation.

References

Arango, L., Singaraju, S. P., and Niininen, O. (2023). Consumer responses to AI-generated charitable giving ads. Journal of Advertising, 52(4), 486-503. https://doi.org/10.1080/00913367.2023.2183285

Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. Journal of Computer-Mediated Communication, 3(2).

Boerman, S. C., Willemsen, L. M., and Van Der Aa, E. P. (2017). 'This post is sponsored': Effects of sponsorship disclosure on persuasion knowledge and electronic word of mouth in the context of Facebook. Journal of Interactive Marketing, 38, 82-92. https://doi.org/10.1016/j.intmar.2016.12.002

Campbell, C., Plangger, K., Sands, S., and Kietzmann, J. (2022). Preparing for an era of deepfakes and AI-generated ads: A framework for understanding responses to manipulated advertising. Journal of Advertising, 51(1), 22-38. https://doi.org/10.1080/00913367.2021.1909515

Campbell, M. C., and Kirmani, A. (2000). Consumers' use of persuasion knowledge: The effects of accessibility and cognitive capacity on perceptions of an influence agent. Journal of Consumer Research, 27(1), 69-83. https://doi.org/10.1086/314309

Castelo, N., Bos, M. W., and Lehmann, D. R. (2019). Task-dependent algorithm aversion. Journal of Marketing Research, 56(5), 809-825. https://doi.org/10.1177/0022243719851788

Darke, P. R., and Ritchie, R. J. B. (2007). The defensive consumer: Advertising deception, defensive processing, and distrust. Journal of Marketing Research, 44(1), 114-127. https://doi.org/10.1509/jmkr.44.1.114

Delgado-Ballester, E. (2004). Applicability of a brand trust scale across product categories: A multigroup invariance analysis. European Journal of Marketing, 38(5/6), 573-592. https://doi.org/10.1108/03090560410529222

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114-126. https://doi.org/10.1037/xge0000033

Epley, N., Waytz, A., and Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864-886. https://doi.org/10.1037/0033-295X.114.4.864

Evans, N. J., and Park, D. (2015). Rethinking the persuasion knowledge model: Schematic antecedents and associative outcomes of persuasion knowledge activation for covert advertising. Journal of Current Issues and Research in Advertising, 36(2), 157-176. https://doi.org/10.1080/10641734.2015.1023873

Friestad, M., and Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. Journal of Consumer Research, 21(1), 1-31. https://doi.org/10.1086/209380

Glikson, E., and Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627-660. https://doi.org/10.5465/annals.2018.0057

Go, E., and Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in Human Behavior, 97, 304-316. https://doi.org/10.1016/j.chb.2019.01.020

Hayes, A. F. (2022). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (3rd ed.). The Guilford Press.

Huh, J., Nelson, M. R., and Russell, C. A. (2023). ChatGPT, AI advertising, and advertising research and education. Journal of Advertising, 52(4), 477-482. https://doi.org/10.1080/00913367.2023.2227013

Lankton, N. K., McKnight, D. H., and Tripp, J. F. (2015). Technology, humanness, and trust: Rethinking trust in technology. Journal of the Association for Information Systems, 16(10), 880-918. https://doi.org/10.17705/1jais.00411

Long, D., and Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

MacKenzie, S. B., and Lutz, R. J. (1989). An empirical examination of the structural antecedents of attitude toward the ad in an advertising pretesting context. Journal of Marketing, 53(2), 48-65. https://doi.org/10.1177/002224298905300204

Mathur, M. B., and Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. Cognition, 146, 22-32. https://doi.org/10.1016/j.cognition.2015.09.008

Puntoni, S., Reczek, R. W., Giesler, M., and Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. Journal of Marketing, 85(1), 131-151. https://doi.org/10.1177/0022242920953847

Reeves, B., and Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people and places. CSLI Publications.

Short, J., Williams, E., and Christie, B. (1976). The social psychology of telecommunications. John Wiley & Sons.

Soh, H., Reid, L. N., and King, K. W. (2009). Measuring trust in advertising. Journal of Advertising, 38(2), 83-104. https://doi.org/10.2753/JOA0091-3367-38.2.83

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

Tu, C. H. (2002). The measurement of social presence in an online learning environment. International Journal on E-Learning, 1(2), 34-45.

Van Noort, G., Himelboim, I., Martin, J., and Collinger, T. (2020). Introducing a model of automated brand-generated content in an era of computational advertising. Journal of Advertising, 49(4), 411-427. https://doi.org/10.1080/00913367.2020.1795954

Wortel, C., Vanwesenbeeck, I., and Tomas, F. (2024). Made with artificial intelligence: The effect of artificial intelligence disclosures in Instagram advertisements on consumer attitudes.

Journal of Current Issues & Research in Advertising, 2(3).
https://doi.org/10.1177/27523543241292096

Wu, L., and Wen, T. J. (2021). Understanding AI advertising from the consumer perspective. Journal of Advertising Research, 61(2), 133-146. https://doi.org/10.2501/JAR-2021-004

Ethics Declarations

Ethical approval. The study was conducted in South Korea as an anonymous, minimal-risk online survey experiment involving adult participants. No formal ethics approval or formal waiver was issued by an institutional review board or research ethics committee. The study was assessed by the authors as meeting the exemption criteria for minimal-risk human-subjects research under the Republic of Korea Bioethics and Safety Act, Article 15(2), and the Enforcement Rule of the Bioethics and Safety Act, Article 13(1), which provide that certain human-subjects research may be exempt from institutional review when it uses publicly available information or does not collect or record personally identifiable information and meets the specified minimal-risk categories.[1] The basis for applying this exemption criterion was that the study involved anonymous responses from adult participants, collected no personally identifiable or sensitive personal information, involved no vulnerable participants, and posed no more than minimal risk. All research procedures were performed in accordance with relevant South Korean regulations governing human-subjects research and the ethical principles of the Declaration of Helsinki, including voluntary participation, informed consent, anonymity, confidentiality of responses, and the right to withdraw from the study.

[1] Official regulation/policy URL: Korean Law Information Center, Bioethics and Safety Act, Article 15(2), and Enforcement Rule of the Bioethics and Safety Act, Article 13(1), Republic of Korea, Ministry of Health and Welfare: https://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=271595

Informed consent. Informed consent was obtained from all participants before taking part in the study.

Data availability

The de-identified dataset, variable definitions and coding scheme, analysis syntax, and supporting research materials used to support the findings of this study have been provided as related files for editorial and peer-review evaluation. Public deposition of the dataset was not included in the original participant consent procedure; therefore, the data are not publicly archived. Additional information is available from the corresponding author upon reasonable request.