Original Article

Journalism & Mass Communication Quarterly
1–22

DOI: 10.1177/10776990261462536
http://journals.sagepub.com/home/jmq
Sage

Perceived Legitimacy Matters: Building Public Trust and Acceptance of AI-Generated News Images Through Strategic AI Disclosure

Yuhan Li[1] **and Hang Lu**[1]

Abstract

Despite growing newsroom interest in generative visual AI, the use of AI-generated images may backfire. An online experiment ($N = 1{,}171$) shows that regardless of disclosure strategies, AI-generated images reduce trust in news organizations and acceptance of AI in journalism compared to real photos. However, attributing AI use to external motives (i.e., protecting victims' privacy) elicits more favorable responses than internal attribution (i.e., saving newsroom costs) or simple disclaimers. Perceived legitimacy underlies audience response, and people's predispositions toward AI moderated the effectiveness of disclosure strategies. These findings extend AI transparency research and help newsrooms to communicate AI use more ethically.

Keywords

AI-generated news images, perceived legitimacy, trust in news organizations, acceptance of AI in journalism, strategic AI disclosure, journalistic innovation

Introduction

In many high-stakes and dangerous reporting situations, from active conflict zones to natural disasters, newsrooms often face significant challenges in accessing the scene

[1]University of Michigan, Ann Arbor, USA

Corresponding Authors:
Yuhan Li, University of Michigan, 105 S. State Street, Ann Arbor, MI 48109-1079, USA.
Email: yhanli@umich.edu

Hang Lu, University of Michigan, 105 S State Street, Ann Arbor, MI 48109, USA.
Email: hanglu@umich.edu

and obtaining timely visuals. These environments are often physically inaccessible due to safety concerns, transportation shutdowns, and disrupted communication. This study examines disaster news reporting as an important case of this broader problem: when disasters such as volcanic eruptions or avalanches strike, newsrooms may struggle to document unfolding crises in real time. Yet, in the early stages of natural disasters, visuals play an important role in communicating risks, sustaining public attention, and mobilizing action (Zelizer, 2010).

Generative visual AI, such as Dall·E 3, Midjourney, and Adobe FireFly, emerges as a potential solution for visualizing events that couldn't be safely or easily photographed (Thomson et al., 2025). While many journalists view this as an innovative solution, the integration of generative AI in newsrooms raises significant ethical concerns and triggers audience backlashes. Indeed, most newsrooms refrain from using AI to create news visuals; When they do publish AI-generated images, they usually note clearly why they did so and how they are made, and apply a label to the image (New York Times, 2024).

While recent studies suggest that audiences show concerns and negative reactions to AI-generated news visuals (Morosoli et al., 2025; Strikovic & Cools, 2025), little is known about whether and how strategic disclosures can attenuate these negative audience reactions. Rather than endorsing the use of AI-generated news images, this study examines how disclosure strategies might shift audience responses. This mirrors broader communication research exploring how transparency strategies shape audience responses to controversial technology use.

Earlier scholarship often sees transparency, a core authoritative ritual in journalism, as a remedy to avoid potential audience backlash to AI-assisted journalism (Diakopoulos & Koliska, 2017). However, recent studies suggest that disclosure does not always improve audience evaluations and may instead reduce perceived believability and trust in news organizations (Toff & Simon, 2024). As AI becomes more deeply integrated in news production, newsrooms face a significant dilemma: the ethical imperative to disclose AI usage versus the potential reputational risks that transparency may incur. This raises a critical question for newsrooms:

How can newsrooms better communicate AI use to the public?

To address this question, we examine how *strategic AI disclosure* shapes perceived legitimacy of using AI-generated news images, trust in the news organization, and acceptance of AI in journalism. Building upon attribution theory (Weiner, 2006), we define strategic AI disclosure as providing either external or internal attributed motives for using AI-generated news images. We compare these disclosure strategies with a simple disclaimer (the long-standing form of AI disclosure) and with a real photo as the baseline. Using an online experiment with mock disaster news, we test how AI-generated images accompanied by different disclosure messages influence audience trust in news organizations and acceptance of AI in journalism.

This study makes both theoretical and practical contributions. First, it advances growing research on AI transparency by moving beyond a binary view of

AI disclosure and testing whether internal and external attributional framing shifts audience responses. Second, it identifies perceived legitimacy as a key psychological mechanism linking AI disclosure strategies to audience responses. Third, by identifying predispositions toward AI as a moderator, it delineates the individual-level boundary conditions that shape audience reception of AI-generated news images. Together, these findings offer practical guidance for newsrooms seeking to strategically integrate AI into news production without compromising public trust.

Literature Review

The Transparency Dilemma and Disclosure Strategies of AI Use

While generative AI has emerged as a useful tool for visual storytelling in time-sensitive and resource-constrained contexts, it raises concerns about misinformation and journalistic ethics (Dan et al., 2021; von Sikorski & Hameleers, 2025). Transparency is often seen as a remedy to navigate the tension between misinformation and responsible AI-assisted journalism (Diakopoulos & Koliska, 2017). Despite a lack of formal guidelines, long-standing journalistic norms emphasize the need for transparent and responsible AI use in newsrooms (Karlsson, 2010).

However, transparency alone does not necessarily resolve this tension. In some cases, disclosing AI use can even backfire. For instance, Toff and Simon (2024) found that audiences consistently rated AI-labeled news articles as less credible. Similarly, Wittenberg et al. (2025) found that labeling AI-generated images significantly reduced the perceived credibility of presented content, regardless of what types of labels (i.e., explaining how content was made or highlighting content's potential to mislead) were used. In advertising contexts, Baek et al. (2026) found that disclosing AI-generated images significantly reduced credibility perceptions and led to unfavorable attitudes toward advertisements. These findings point to a disclosure dilemma: although transparency may fulfill ethical standards, it can also inadvertently erode public trust.

In response, scholars call for more strategic disclosure strategies that move beyond treating "disclosure" as a binary (i.e., either with a disclaimer of "generated by AI" or absent) and consider how the framing of disclosure shapes audience perceptions (K. de Fine Licht & J. de Fine Licht, 2020). Rather than advocating for a "full transparency" approach, K. de Fine Licht and J. de Fine Licht (2020) propose that providing justifications for the decision to use AI (i.e., transparency in rationale) can reduce public backlash. That is, disclosure should not merely state that AI was used but also provide context and justification for its use. Strategic disclosure, particularly when it includes clear rationales, can enhance public understanding, mitigate skepticism, and potentially build trust (J. de Fine Licht et al., 2014).

Building on this perspective, we focus on one specific journalistic application: the use of AI-generated news images in disaster reporting. As a major component of news production, visuals help convey the immediacy and urgency of disasters and activate public attention (Zelizer, 2010). However, news organizations show concerns about using generative visual AI, such as potential mis/disinformation and reputational risk

(Thomson et al., 2025). To address this, our study seeks to examine whether attributed motives for using AI-generated news images influence audience responses, offering insights into how newsrooms can better communicate AI use and mitigate potential backlash.

Drawing on attribution theory (Weiner, 2006), we conceptualize strategic AI disclosure in terms of internal versus external attributions. Attribution theory posits that individuals have a basic psychological need to understand the causes of various issues and events as a way to adjust to and control their environment (Heider, 1958). People tend to attribute events to either internal factors (e.g., egoistic or self-serving motives) or external causes (e.g., altruistic motives) (Heider, 1958; Jones & Davis, 1965), among which factors such as assumed intentionality, perceived locus of control, and social desirability of the behavior influence the types of attributions made (Ramasubramanian, 2011; Weiner, 1986). In journalistic contexts, news stories often use framing strategies that imply causes and consequences to help their audiences make sense of various social issues and shape their subsequent reactions and attitudes (Coombs, 2007; Iyengar, 1990; Ramasubramanian, 2011). For example, Lu (2024) found that news coverage of the opioid crisis using external attributions increased public support for perpetrator-oriented punitive policies.

Although prior research focused mainly on how news frames shape responsibility attribution, AI disclosure is essentially a similar message that utilizes framing devices to imply motivations and causes of the social entity. Different framing of AI disclosure messages can provide subtle cues that facilitate certain causal interpretations and responsibility attributions (Entman, 1993; Ji et al., 2025). Disclosures that emphasize self-serving motives foreground internal causes, whereas disclosures that stress altruistic motives and social benefits highlight external causes. Prior research suggests that when negative outcomes are attributed to external motives, actors are less blamed. When the same outcomes are attributed to internal motives, they are seen as less justified and may receive more blame (Jeong, 2009; Weiner, 2006).

Audiences may apply similar explanatory judgment processes in an attempt to evaluate AI use in the journalistic context (K. de Fine Licht & J. de Fine Licht, 2020). Justifying AI use from external causes, such as protecting victims' privacy or illustrating inaccessible disaster scenes, is more likely to be interpreted as externally motivated and ethically grounded. Such rationales align with professional standards of journalism and may buffer against the trust penalty of using AI in news (Karlsson, 2010). In contrast, justifying AI use from internal causes, such as reducing the newsroom's travel costs and accelerating news production, may appear internally motivated and trigger punitive attitudes. Importantly, our design also includes a control condition featuring a real photograph, allowing us to compare legitimacy perceptions of AI-generated images under different disclosure messages against a baseline of traditional photojournalism.

Based on the above, this study investigates how disclosure strategies influence both acceptance of AI in journalism and trust in the news organization. Acceptance of AI in journalism reflects broader audience attitudes and has important implications for predicting support for AI integration in future journalism (Kelly et al., 2023). Understanding

how disclosure strategies shape AI acceptance offers important insights into the public's openness to technological innovation in journalism. Trust in news organizations is also critical, given evidence that disclosing AI use reduced perceived trustworthiness of news content and undermined audience confidence in journalism as an institution (Toff and Simon, 2024). Considering both outcomes allows us to provide a more holistic view of how audiences evaluate AI-generated visuals in journalistic contexts.

Given the above, we hypothesize that framing AI use with an external attribution will yield more positive audience responses, compared to a simple disclaimer (**H1**) and an internal attribution (**H2**):

H1: Compared to a simple AI disclaimer, an external attribution for using AI-generated images will increase (a) trust in news organizations and (b) acceptance of AI in journalism.

H2: Compared to an internal attribution, an external attribution for using AI-generated images will increase (a) trust in news organizations and (b) acceptance of AI in journalism.

Perceived Legitimacy as the Key Mechanism in AI Disclosure Effects

As news organizations increasingly adopt AI into news production and dissemination, its use in journalism has also resulted in audience backlashes, including reduced public trust, heightened risk perceptions, and lower acceptance of AI in future journalism (Mitova et al., 2025; Morosoli et al., 2025). We argue that perceived legitimacy is a key mechanism linking AI disclosure to these negative outcomes. Legitimacy refers to a generalized perception that an entity's actions "are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (Suchman, 1995). In journalism, perceived legitimacy depends on whether audiences believe newsroom practices align with professional and ethical norms (Tong, 2018). People's concerns about legitimacy would arise when news organizations take unconventional actions that deviate from established norms (Schilke & Reimann, 2025). For instance, Blassnig et al. (2024) found that when audiences perceive editorial gatekeeping as being replaced by invisible news recommender systems, trust in those outlets declines, as it changes the traditional way media and audiences interact. Similarly, in the context of AI-generated news images, audiences may question the newsroom's legitimacy if they view such practices as violating journalistic standards of authenticity and ethics.

Within attribution theory, disclosure strategies that imply external or internal attribution motives for AI use can shape perceived legitimacy of such behaviors (Jahn et al., 2020). External attributions that highlight altruistic motives are more likely to be seen as socially responsible, thereby enhancing perceived legitimacy (Jahn et al., 2020). In contrast, internal attributions that signal self-serving motives are likely to weaken perceived legitimacy (Coombs, 2007). Prior organizational research supports this logic. Miotto and Youn (2020) found that altruistic attributions of a company's

sustainable collections increased perceptions of corporate legitimacy, whereas Jahn et al. (2020) showed that reward-driven motives reduced both corporate credibility and organizational legitimacy.

Applied to AI-assisted journalism, we argue that when audiences perceive journalistic AI use as driven primarily by expedience or self-serving goals, they are more likely to form internal attributions and reduce perceived legitimacy. When AI usage is framed as serving public rights or the stakeholder interests, they are more likely to form external attributions, thereby enhancing perceived legitimacy. A simple AI disclaimer (e.g., "This image was generated by AI") is likely to perform worse than an external attribution because it provides no contextual rationale for AI use. Without an explicit explanation, audiences must infer the newsroom's motives themselves, often relying on their preexisting dispositions toward AI. Given that prior research has shown that citizens are likely to show resistance to integrating AI into the news cycle (Morosoli et al., 2025), a simple AI disclaimer without any rationales given will result in more negative judgments. The absence of a rationale leaves room for audiences to assume convenience-driven motives and conclude that the newsroom might just choose AI for expedience, which could lower perceived legitimacy.

Taken together, we expect that disclosure strategies framing AI use as externally motivated are more likely to enhance perceived legitimacy than strategies that offer no rationale or highlight internal attributions.

H3: Compared to a simple AI disclaimer, an external attribution will increase perceived legitimacy of using AI-generated images in news.

H4: Compared to an internal attribution, an external attribution will increase perceived legitimacy of using AI-generated images in news.

Perceived legitimacy, in turn, should shape public trust in news organizations and the broader use of AI in journalism. According to organizational legitimacy theory (Deephouse & Suchman, 2008), when audiences evaluate an organization as legitimate and perceive it as aligned with social norms and ethical standards, they would support and trust that organization; when they view it as illegitimate, audiences would question its integrity and reliability, leading to distrust or other negative reactions. Mitova et al. (2025) found that citizens' trust in the journalistic deployment of AI-powered tools remains relatively low. Similarly, Toff and Simon (2024) found that audiences perceived news labeled as AI-generated as less trustworthy, even when articles themselves are not evaluated as any less accurate. This suggests that people's distrust and increased scrutiny of AI-assisted journalism may not only concern the accuracy of AI-generated content but also other factors, such as the ethical standards of AI-assisted news production (Culver & Lee, 2019). We therefore argue that reduced perceived legitimacy might undermine trust in the news organization.

The same logic applies to the acceptance of AI in journalism, which reflects audiences' willingness to support integrating AI into news production. Acceptance of AI in journalism depends not only on perceived usefulness but also on whether the

technology is viewed as normatively appropriate (Martin & Waldman, 2023). When AI use is seen as legitimate, audiences should be more willing to accept its role in journalism. Therefore, we propose:

H5: Perceived legitimacy of the image used in news, whether AI-generated or real, will be positively related to (a) trust in news organizations and (b) acceptance of AI in journalism.

In fact, the mediating role of perceived legitimacy has been found in extant studies on public responses toward AI usage in various contexts. For instance, Schilke and Reimann (2025) found that disclosing AI use to the public could reduce public trust across multiple tasks, such as communication, analytics, and artistry, and the trust erosion was largely explained by damaged legitimacy perceptions. Applying this to AI-assisted journalism, disclosure strategies influence legitimacy perceptions, which in turn shape both public trust in the news organization and acceptance of AI in journalism.

H6: Perceived legitimacy of the image used in news, whether AI-generated or real, will mediate the relationship between image exposure and (a) trust in news organizations and (b) acceptance of AI in journalism.

How Attitudes Toward AI Moderate the Effects of Disclosure Strategies

While strategic disclosure is designed to foster public trust and enhance acceptance of AI-generated images in journalism by increasing the perceived legitimacy of AI use, audiences may not uniformly accept the rationales offered by news organizations (Morosoli et al., 2025). We argue that the impact of disclosure strategies is likely contingent on individuals' underlying attitudes toward AI. Attitudes toward AI, ranging from techno-optimism to skepticism or technophobia, can fundamentally shape how audiences interpret disclosure messages. For instance, individuals with more positive attitudes often view AI adoption as innovative and efficient (Bewersdorff et al., 2025; Morosoli et al., 2025). They may therefore respond more favorably when disclosures frame AI use in socially responsible terms (e.g., external attributions) but still become skeptical when AI is presented as self-serving or when disclosure is minimal (e.g., a simple AI disclaimer). In contrast, individuals with negative attitudes toward AI are predisposed to perceive AI involvement as risky, threatening, or deceptive. For them, even socially responsible framings may not fully mitigate suspicion, and self-serving or minimal disclosures may further erode legitimacy.

These diverging expectations highlight that disclosure strategies are unlikely to affect all audiences in the same way. Yet, existing research has not examined whether and how attitudes toward AI interact with disclosure strategies to shape perceived legitimacy, which subsequently influences trust and AI acceptance. To address this gap, we propose the following research question and present our conceptual model in Figure 1:

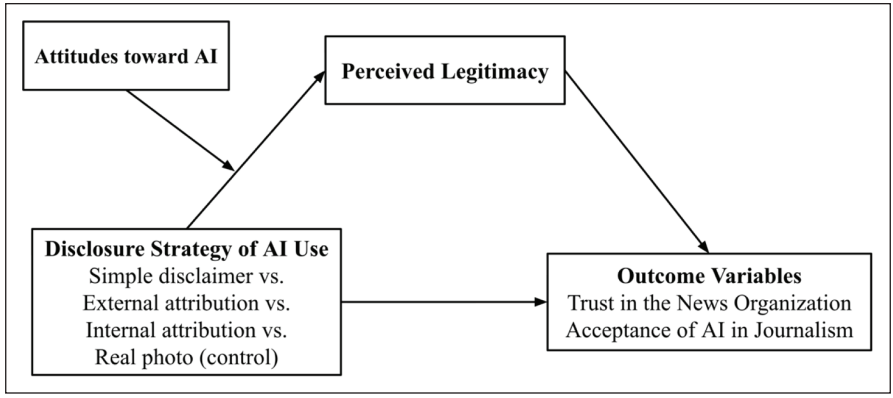

Figure 1. The conceptual model of this study.

RQ1. Do people's attitudes toward AI moderate the indirect effects of disclosure strategies on trust in news organizations and acceptance of AI in journalism via perceived legitimacy?

Methods and Data

Overview of the Study

This study was approved by the Institutional Review Boards at the University of Michigan. To examine the proposed hypotheses and research questions, we conducted a 4 (Disclosure Strategy: Simple AI Disclaimer vs. Internal Attribution vs. External Attribution vs. Real Photo) × 2 (Victim Count: Multiple vs. Single) × 2 (Disaster Context: Volcano Eruption vs. Avalanche) between-subjects experiment.

Participants first completed a pretreatment survey measuring their attitudes toward AI. They were then shown a mock breaking news article attributed to CBS News, describing either a recent avalanche or volcanic eruption and its associated victims. The two disaster issues were used as a form of stimulus sampling, which improves construct and external validity (Wells & Windschitl, 1999). Based on the identifiable victim effect (Slovic, 2007), the number of victims depicted in the AI-generated news image was manipulated to show either a single victim or five victims, and we aimed to examine whether the number of victims depicted in AI-generated disaster news images would influence moral judgments of AI use. However, victim count did not produce significant main effects or interactions with disclosure strategy on any outcome variable (all $ps > .05$). Given space constraints, we do not report these results in the main text. Means and standard deviations for all experimental conditions, including victim count, are reported in the Supplemental Materials.

Participants were randomly assigned to see versions of the news article that varied along the following dimensions: (a) disclosure strategy: the article included either a real photo (no AI disclosure), an AI-generated image with a simple disclaimer (no

justification), an AI-generated image attributed to an internal reason (i.e., cost-saving), or an AI-generated image attributed to an external reason (i.e., protecting the victim's privacy); (b) Victim count: the article depicted either one or five victims; and (c) Disaster context: the article described either a volcanic eruption or an avalanche. Immediately after reading their assigned article, participants completed a post-treatment survey measuring their attitudes toward the use of AI-generated images in the news article. At the end of the survey, all participants were debriefed and informed about the purpose of this study and the experimental manipulations.

Design and Stimuli

We created the stimuli in May 2025, comprising three elements: news articles, AI-generated images, and disclosure strategies. The articles were adapted from two real news articles about volcanic eruptions and avalanches, but were modified to reflect breaking news scenarios where firsthand reporting was infeasible. Each original article was rewritten into a 110-word news story with fabricated details (e.g., author, date, and location) to fit our research context. To manipulate the victim count, we created two versions of each article, one describing a single victim and the other describing multiple victims, while keeping all other content consistent.

We utilized Sora, a text-to-image model developed by OpenAI, to generate news images tailored to the news article. The first author iteratively crafted and refined prompts based on each article to produce realistic disaster images. When outputs were unsatisfactory or contained any flaws, the author revised and resubmitted the prompts. This process was repeated until high-quality images were obtained. All article versions, image generation prompts, and images were included in the Supplemental Materials.

To develop appropriate AI disclosure strategies, we conducted a pilot study on June 1, 2025, to identify internal and external attributions that align with public perceptions. Specifically, we sought to determine which reasons for using AI-generated images in disaster news were perceived as altruistic (external) or self-serving (internal). We began by drafting five candidate justifications. Then, we recruited 100 participants via Prolific and randomly assigned them to rate the perceived self-serving intentions of each justification. The results suggest that the justification, protecting the victims' privacy, was perceived as the least self-serving ($M=2.36$) justification, whereas cost-saving was rated as the most self-serving ($M=4.03$) reason. Paired t-test confirmed these differences were statistically significant ($t=-10.05$, $p<.001$). Thus, we selected "protecting the victim's privacy" as external attribution and "cost-saving" as internal attribution. Figure 2 illustrates the sample news article and accompanying images, each with different disclosure messages. More details about the pilot study and selection process are included in the Supplemental Materials.

Participants

We conducted the main study on June 5, 2025, recruiting a total of 1,381 English-speaking participants based in the United States from Prolific. We chose Prolific

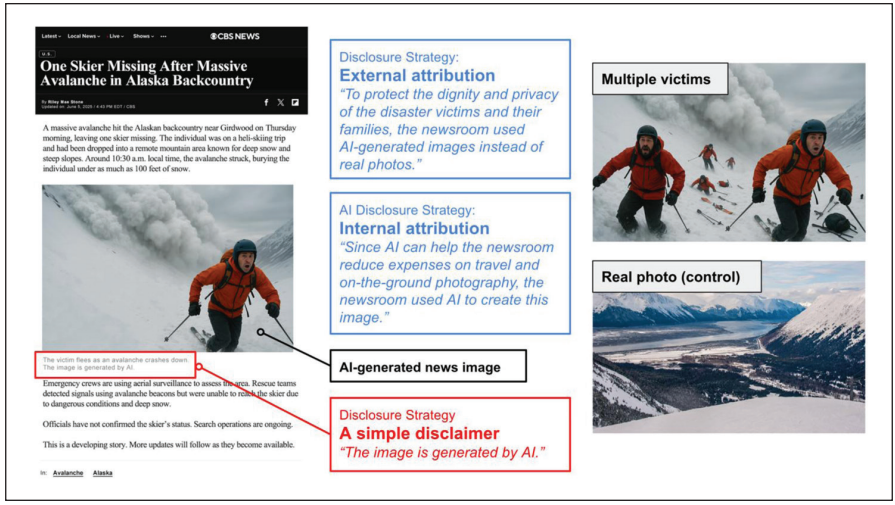

Figure 2. The avalanche disaster news article included an AI-generated image accompanied by different disclosure strategies. The versions featuring multiple victims or the real photo (control) are shown on the right side. See Supplemental Materials for all stimuli.

because it can provide comparatively high-quality data, including stronger attention and comprehension performance than other crowdsourced platforms (Peer et al., 2021). Each participant received an average compensation of $2.5. Participants were randomly assigned to one of the sixteen experimental conditions described previously. Individuals who had participated in the pilot study were not eligible for the main study. To ensure data quality, we applied multiple exclusion criteria. First, we removed participants who either failed at least one of two attention check questions or reported issues during the survey, resulting in 1,298 valid responses. Next, we excluded participants with unusually long survey completion times, since unusual long durations could indicate that respondents became distracted or disengaged partway through the study. Following the outlier detection approach proposed by Leys et al. (2013), we excluded cases with completion times exceeding three median absolute deviations (MADs) above the median. This resulted in a final sample of 1,171 participants.

Among the participants, 45.35% identified as male and 52.86% identified as female, and 1.79% preferred not to disclose. The average age of participants was 40.13 ($SD = 13.98$). The median education level was a bachelor's degree, and the median annual household income fell within $70,000 to $79,999. The racial or ethnic background of participants was as follows: White 63.88%, Black 24.77%, Asian 4.01%, Hispanic or Latine 3.84%, American Indian or Alaska Native 0.85%, Native Hawaiian or other Pacific Islander 0.17%, Other racial or ethnic types 1.79%, and 0.68% preferred not to disclose.

Measures

All items were measured using a 5-point scale ranging from 1 = strongly disagree to 5 = strongly agree. The mean and standard deviation for each condition, full-scale items, and correlation plots for all measured variables are provided in the Supplemental Materials.

Attitudes Toward AI. To measure people's attitudes toward AI, we adapted four items from Grassini (2023) and asked participants to indicate their agreement with statements such as: "I am excited about the potential benefits of AI," "AI will improve our quality of life," and "AI is likely to do more harm than good in the long run" (reverse-coded) ($\alpha = .92$, $M = 2.76$, $SD = 1.27$).

Perceived Legitimacy. Participants indicated their agreement with the following seven statements that were adapted from Martin and Waldman (2023): "I think the use of the image in the news article is legitimate / justifiable / appropriate / ethical / reasonable / complying with relevant policies and laws / aligning with professional journalistic norms" ($\alpha = .97$, $M = 3.16$, $SD = 1.21$).

Acceptance of AI in Journalism. To assess public acceptance of AI-generated visuals in journalism, we asked participants for their agreement with four statements, such as: "It is acceptable to use AI to illustrate disasters in news coverage," "It is acceptable to use AI to depict disaster victims in news coverage" ($\alpha = .92$, $M = 2.76$, $SD = 1.27$).

Trust in the News Organization. We asked a question adapted from Toff and Simon (2024) to capture public trust in the news organization: "To what extent do you find the news organization that published this article trustworthy?" ($M = 2.87$, $SD = 1.09$).

Analytic Approach

To examine **H1**–**H4**, which predicted main effects of the disclosure strategy condition, we conducted three-way ANOVAs with disclosure strategy, victim count, disaster context, and their two-way and three-way interactions as the independent variables, and trust in the news organization, acceptance of AI in journalism, and perceived legitimacy as the dependent variables.

For **H5**, we conducted two hierarchical regression analyses. Models 1 and 2 predicted trust in the news organization, whereas Models 3 and 4 predicted acceptance of AI in journalism. The experimental conditions were entered in Model 1 and Model 3. In Model 2 and Model 4, we added perceived legitimacy of AI-generated image use to assess whether it predicted additional variance in the outcome variables beyond the experimental manipulations. To examine **H6**, which concerned mediation effects, we used the PROCESS macro Model 4 in SPSS with multi-categorical independent variables (dummy-coded), 5,000 bootstrap resamples, and 95% confidence intervals (Hayes, 2017). We conducted mediation analyses for six pairwise contrasts: simple disclaimer versus control; external attribution versus control; internal attribution

versus control; simple disclaimer versus external attribution; simple disclaimer versus internal attribution; external attribution versus internal attribution. In each model, the focal contrast was the independent variable, perceived legitimacy was the mediator, and either trust or AI acceptance was the dependent variable.

To answer **RQ1**, we used PROCESS Model 7 with the same multi-categorical independent variables, bootstrap parameters, and pairwise contrasts. Attitudes toward AI were modeled as a continuous moderator of the indirect effects of disclosure strategies (independent variable) on trust or AI acceptance (dependent variable) through perceived legitimacy (mediator). Given that the three-way ANOVAs yielded no significant two- or three-way interactions, we opted to exclude these terms to avoid unnecessary model complexity and to maintain parsimony. Consequently, only disaster context and victim count were incorporated as covariates in the mediation and moderated mediation analyses.

Results

Manipulation Checks

For the manipulation of disclosure strategies, we asked participants exposed to either external or internal attribution about their perceptions of the newsroom's motive for using an AI-generated image (1 = entirely altruistic, 5 = entirely self-serving). For the victim count, participants were asked whether the news image included at least one person (0 = no, 1 = yes), and if they said yes, we further asked them to indicate whether the image showed more than one person (0 = no, 1 = yes). For the manipulation of disaster context, all participants identified the disaster described in the article (1 = Volcanic Eruption, 2 = Hurricane, 3 = Wildfire, 4 = Avalanche).

The one-way ANOVAs results showed that participants in the internal attribution condition were significantly more likely to perceive the justification as self-serving ($M = 3.46$, $SD = 1.22$) than those in the external attribution condition ($M = 3.12$, $SD = 1.23$), $F(1, 592) = 11.22$, $p < .001$, partial $\eta^2 = .019$, suggesting our manipulation of disclosure strategies was successful. Moreover, our chi-square tests showed that 95.4% of participants in the multiple-victim condition reported seeing more than one person, while 95.4% in the single-victim condition reported seeing only one, $\chi^2(1) = 751.73$, $\phi = .910$, $p < .001$, which validates our manipulation of the victim count. For the disaster context, our chi-square results indicated that most participants in the volcano condition identified it as a volcanic eruption (98.9%), and most in the avalanche condition selected avalanche (99.3%), $\chi^2(3) = 1{,}099.93$, $p < .001$, suggesting this manipulation was also successful.

Effects of Disclosure Strategy on Trust, AI Acceptance, and Perceived Legitimacy

Our **H1−H4** proposed that compared to the simple disclaimer and internal attribution, external attribution would increase trust, AI acceptance, and perceived legitimacy. Indeed, there was a significant main effect of the disclosure strategy condition on trust in the news organization, $F(3, 1{,}155) = 24.58$, $p < .001$, $\eta^2 = .06$. Post hoc pairwise

comparisons showed that participants who viewed real photos (control condition) reported significantly higher trust ($M = 3.34$, $SD = 0.89$) than those shown AI-generated images with a simple disclaimer ($M = 2.67$, $SD = 1.15$; $p < .001$), an external attribution ($M = 2.82$, $SD = 1.11$; $p < .001$), and internal attribution ($M = 2.67$, $SD = 1.08$; $p < .001$). However, trust levels did not significantly differ among the three AI disclosure conditions, providing no support for **H1a** and **H2a**. In summary, we found that regardless of how AI usage was disclosed, AI-generated images consistently reduced public trust in news organizations.

We found a similar pattern for acceptance of AI in journalism. There was a significant main effect of disclosure strategy, $F(3, 1{,}155) = 3.66$, $p = .012$, $\eta^2 = .009$. Participants in the real photo condition reported greater AI acceptance ($M = 2.93$, $SD = 1.21$) than those in the simple disclaimer condition ($M = 2.59$, $SD = 1.27$; $p < .001$). However, AI acceptance did not differ significantly between the control and external attribution ($M = 2.81$, $SD = 1.30$; $p = .264$), or between the control and internal attribution ($M = 2.74$, $SD = 1.27$; $p = .071$). Participants in the external attribution condition reported significantly higher AI acceptance than those in the simple disclaimer condition ($p = .035$), while there were no significant differences between the simple disclaimer and internal attribution ($p = .143$) or between the external and internal attribution ($p = .498$). **H1b** got supported, but **H2b** was rejected.

We found a significant main effect of disclosure strategy on perceived legitimacy, $F(3, 1167) = 73.49$, $p < .001$, $\eta^2 = .160$. Participants in the control (real photo) condition reported the highest perceived legitimacy ($M = 4.00$, $SD = 0.73$), significantly higher than those in the simple disclaimer ($M = 2.78$, $SD = 1.20$; $p < .001$), external attribution ($M = 3.05$, $SD = 1.21$; $p < .001$), and internal attribution ($M = 2.86$, $SD = 1.19$; $p < .001$) conditions. External attribution was perceived to be more legitimate than the simple disclaimer ($p = .003$) and internal attribution ($p = .030$). Both **H3** and **H4** were supported.

Perceived Legitimacy Mediated the Effects of Disclosure Strategies on Trust and AI Acceptance

To test **H5**, we ran hierarchical regression models. For trust in the news organization, Model 1 (without legitimacy) was significant, $F(5, 1165) = 16.07$, $p < .001$, adjusted $R^2 = .060$. Adding perceived legitimacy in Model 2 significantly improved model fit, $F(6, 1164) = 137.50$, $p < .001$, with adjusted R^2 increasing to .412. Perceived legitimacy strongly predicted trust ($b = 0.588$, $SE = 0.022$, $p < .001$). **H5a** was supported. For AI acceptance, Model 3 was not significant, $F(5, 1165) = 2.22$, $p = .050$, adjusted $R^2 = 0.005$. However, including perceived legitimacy in Model 4 significantly improved the model fit, $F(6, 1164) = 145.8$, $p < .001$, with adjusted R^2 increasing to .426. Perceived legitimacy strongly predicted AI acceptance ($b = 0.746$, $SE = 0.026$, $p < .001$), supporting **H5b**. Full regression results are provided in Table 1.

We then tested **H6**, which predicted that perceived legitimacy would mediate the relationship between disclosure strategies and both trust (**H6a**) and AI acceptance (**H6b**). Using the real photo as the reference, all three AI disclosure strategy conditions produced significant negative indirect effects on trust via perceived legitimacy: simple

Table 1. Hierarchical Regression Results.

	Trust in news organization		Acceptance of AI in journalism	
	Model 1	Model 2	Model 3	Model 4
Disclosure strategy (ref. Real photo)				
Simple disclaimer	−0.661*** (0.089)	0.057 (0.075)	−0.342** (0.106)	0.568*** (0.086)
External attribution	−0.510*** (0.089)	0.044 (0.074)	−0.121 (0.107)	0.582** (0.085)
Internal attribution	−0.665*** (0.088)	0.009 (0.074)	−0.190 (0.105)	0.664*** (0.085)
Disaster: volcano eruption (vs. avalanche)	−0.133* (0.062)	−0.058 (0.049)	0.022 (0.074)	0.117* (0.056)
Single victim (vs. multiple victims)	0.039 (0.062)	−0.019 (0.049)	−0.004 (0.074)	−0.078 (0.056)
Perceived legitimacy		0.588*** (0.022)		0.746*** (0.026)
Constant	3.384*** (0.076)	1.024*** (0.108)	2.922*** (0.090)	−0.073 (0.123)
Observations	1,171	1,171	1,171	1,171
R^2	0.065	0.415	0.009	0.429
Adjusted R^2	0.061	0.412	0.005	0.426

*$p < .05$; **$p < .01$; ***$p < .001$.

disclaimer ($b_{\text{indirect}} = -0.718$, $SE_{\text{indirect}} = 0.058$, 95%CI = [−0.832, −0.602]), external attribution ($b_{\text{indirect}} = -0.554$, $SE_{\text{indirect}} = 0.054$, 95%CI = [−0.659, −0.451]), and internal attribution ($b_{\text{indirect}} = -0.673$, $SE_{\text{indirect}} = 0.056$, 95%CI = [−0.785, −0.566]). When using the simple disclaimer as the reference, the indirect effect of external attribution was significantly positive ($b_{\text{indirect}} = 0.164$, $SE_{\text{indirect}} = 0.059$, 95%CI = [0.048, 0.279]), whereas the indirect effect of internal attribution was non-significant ($b_{\text{indirect}} = 0.044$, $SE_{\text{indirect}} = 0.057$, 95%CI = [−0.066, 0.159]). Using internal attribution as the reference, the indirect effect of external attribution on trust through perceived legitimacy was significantly positive ($b_{\text{indirect}} = 0.119$, $SE_{\text{indirect}} = 0.057$, 95%CI = [0.006, 0.229]).

For AI acceptance (**H6b**), compared to the real photo condition, the simple disclaimer condition ($b_{\text{indirect}} = -0.911$, $SE_{\text{indirect}} = 0.070$, 95%CI = [−1.051, −0.775]), external attribution ($b_{\text{indirect}} = -0.703$, $SE_{\text{indirect}} = 0.069$, 95%CI = [−0.841, −0.570]), and internal attribution ($b_{\text{indirect}} = -0.854$, $SE_{\text{indirect}} = 0.068$, 95%CI = [−0.986, −0.719]) had significant negative indirect effects through perceived legitimacy. Using the simple disclaimer as reference, the indirect effect of external attribution on AI acceptance via perceived legitimacy became significantly positive ($b_{\text{indirect}} = 0.208$, $SE_{\text{indirect}} = 0.073$, 95%CI = [0.062, 0.354]), while the indirect effect of internal attribution was non-significant ($b_{\text{indirect}} = 0.056$, $SE_{\text{indirect}} = 0.073$, 95%CI = [−0.083, 0.200]). Compared to internal attribution, external attribution had a significantly positive indirect effect on AI acceptance via perceived legitimacy ($b_{\text{indirect}} = 0.151$, $SE_{\text{indirect}} = 0.073$, 95%CI = [0.011, 0.300]). Based on the above, **H6** was supported: across comparisons, perceived legitimacy significantly mediated the effects of different disclosure strategies on trust and AI acceptance.

Effects of Disclosure Strategies on Legitimacy Were Conditional on Attitudes Toward AI

Turning to moderated mediation (**RQ1**), we tested whether attitudes toward AI moderated the indirect effects of disclosure strategies on outcomes via perceived legitimacy.

Taking *trust in the news organization* as the outcome and the real photo condition as the reference, the simple disclaimer (index = 0.160, *SE* = 0.050, 95% CI = [0.063, 0.256]), external attribution (index = 0.269, *SE* = 0.047, 95% CI = [0.176, 0.360]), and internal attribution (index = 0.214, *SE* = 0.047, 95% CI = [0.122, 0.304]) showed significant moderated mediation effects on trust. Each AI disclosure condition showed similar patterns across participants with various attitudes toward AI. Specifically, compared to the real photo, all AI disclosures significantly eroded trust through reduced legitimacy, but these indirect effects became weaker among individuals with more positive attitudes toward AI. However, no significant moderated mediation effects were found when comparing different AI disclosure strategies against one another.

We observed a similar pattern when taking the *acceptance of AI in journalism* as the outcome and the real photo condition as the reference. Again, the simple disclaimer (index = 0.202, *SE* = 0.063, 95%CI = [0.080, 0.325]), the external attribution (index = 0.339, *SE* = 0.059, 95%CI = [0.223, 0.449]), and the internal attribution (index = 0.270, *SE* = 0.060, 95%CI = [0.155, 0.387]) produced significant moderated mediation effects. Compared to the real photo, all AI disclosure strategies reduced AI acceptance indirectly via lower legitimacy, but the magnitude of these indirect effects diminished among participants with more positive attitudes toward AI. Further comparisons among different disclosure strategies highlighted the advantage of the external attribution relative to a simple disclaimer (index = 0.137, *SE* = 0.071, 95% CI = [0.003, 0.275]). Compared to the simple disclaimer, the external attribution produced significantly more positive indirect effects on trust among participants with neutral or positive attitudes toward AI, whereas the internal attribution showed no significant moderated mediation effects on trust (index = 0.068, *SE* = 0.073, 95% CI = [−0.073, 0.213]). Using the internal attribution as the reference, the external attribution also showed no significant moderated mediation effects on trust (index = 0.069, *SE* = 0.068, 95% CI = [−0.068, 0.198]). All statistics for the conditional indirect effects, standard errors, and confidence intervals are included in the Supplemental Materials.

Discussion

Building on recent work examining disclosure effects of AI-generated news and imagery (Baek et al., 2026; Strikovic & Cools, 2025; Wittenberg et al., 2025), this study extends the line of inquiry to disaster reporting contexts and offers four main findings. First, compared with real photos, AI-generated news images consistently reduced trust in the news organization and acceptance of AI in journalism. Second, AI disclosure strategies have nuanced implications for audience responses: a simple disclaimer was the least effective approach, whereas external attribution produced more favorable responses than a simple disclaimer, especially by increasing perceived legitimacy and acceptance of AI in journalism. Third, perceived legitimacy emerged as a central mechanism explaining audience negative reactions to AI-generated news images. Fourth, predispositions toward AI moderated the effectiveness of disclosure strategies, and the penalty indirect effects of AI disclosure were attenuated among participants with more positive attitudes toward AI. Together, these findings suggest that strategic disclosure can partially mitigate audience backlash to the use of AI-generated news imagery for certain groups.

Consistent with prior research on negative reactions to AI-generated news imagery (Strikovic & Cools, 2025), our results show that AI-generated visuals significantly decreased both trust in the news organization and acceptance of AI in journalism. Contrary to **H1a** and **H2a**, trust levels did not significantly differ among disclosure strategies, showing that attribution framing cannot restore the trust deficit caused by AI image use. Despite generative visual AI raising opportunities for the production and consumption of visual storytelling (Martinez et al., 2025; Thomson et al., 2025), audiences appear unprepared for its integration into news reporting. To differentiate from misinformation, staff and editors suggested labeling the image source in the accompanying caption (Thomson et al., 2025). However, we found that transparency alone does not necessarily build trust. In fact, a simple AI disclaimer, the most common practice in current journalism, was the least legitimate approach to disclosing AI use to the public.

The picture was more nuanced for the acceptance of AI in journalism and perceived legitimacy. For AI acceptance, external attribution outperformed the simple disclaimer (supporting **H1b**), suggesting that framing AI use as altruistically motivated can at least partially attenuate the penalty effects of AI image use on acceptance of AI in journalism. However, external and internal attribution did not significantly differ from each other (**H2b** not supported), indicating that the advantage of external attribution is specific to its comparison with a simple disclaimer rather than a general superiority over all disclosure strategies. Regarding perceived legitimacy, both **H3** and **H4** were supported. External attribution significantly enhanced perceived legitimacy more than both a simple disclaimer and internal attribution. It might be because an external attribution, such as protecting victims' privacy and dignity in disaster reporting, aligns with journalistic ethics (Culver & Lee, 2019), leading audiences to view such AI use as acceptable. Perceived legitimacy of AI use is important for the debate on what kind of journalistic AI media should strive for. A normative vision of journalistic AI requires more than transparency, but also justifying why AI should be used and aligning with professional criteria (Helberger et al., 2022).

As both **H5** and **H6** were supported, our study further underscores the importance of perceived legitimacy as an explanatory mechanism for audience responses. Legitimacy is central to journalism's authority in the digital era (Carlson, 2017). Audiences evaluate not only whether AI-generated news content is factually correct but also whether the motive behind its use appears responsible and ethically justified (J. de Fine Licht et al., 2014). Our findings found that the damaged trust and acceptance of AI-generated news images (vs. real photos) were significantly mediated by perceived legitimacy. Moreover, external attribution produced more positive indirect effects on both trust and AI acceptance through increased perceived legitimacy than either a simple disclaimer or internal attribution did. These findings also help explain the puzzle identified by Toff and Simon (2024), who found that audiences viewed AI-generated news as less trustworthy even when they did not perceive it as less accurate. Going beyond content accuracy alone, our study suggests that perceived legitimacy is another core pathway through which audiences translate disclosure messages into judgments about AI-assisted journalism practices.

Addressing **RQ1**, we found more nuances among different sub-populations by looking at the interaction between disclosure strategies and attitudes toward AI. Although all AI disclosures lowered trust and AI acceptance compared to real photos, the magnitude of these indirect effects weakened among participants with more positive attitudes toward AI. The moderating role of attitudes toward AI aligns with recent distinctions between "AI advocates" and "cautious critics" (Bewersdorff et al., 2025). For those with more favorable views of AI, using AI-generated images in news may seem less threatening, and they were likely to show fewer negative judgments toward AI-assisted journalism than AI skeptics.

Notably, we found that among participants with more positive attitudes toward AI, external attribution significantly boosted public trust in the news organization via legitimacy, compared to a simple disclaimer. It aligns with Morosoli et al. (2025), who found that the more benefits citizens see in using AI for journalism, the higher their trust is in AI. This study extends this insight by showing that for people who see AI positively, framing the use of AI in an altruistic and prosocial way can activate those positive predispositions and translate them into stronger trust in news organizations. However, for AI skeptics, even an altruistic framing was insufficient to offset their concerns. Future studies should not only consider how AI use is communicated but also how different audiences process these disclosure strategies. Positive attitudes toward AI may boost the effectiveness of external attribution and altruistic framing of AI usage, while negative predispositions may intensify ethical concerns regardless of disclosure strategies.

Theoretically speaking, this study advances research on AI-assisted journalism and AI transparency in several ways. First, it contributes to the burgeoning literature on the consequences of using AI-generated images in journalism and challenges the prevailing assumptions that the effect of transparency is straightforwardly positive. Second, by combining attribution theory with journalism scholarship on legitimacy, we move beyond a binary view of AI disclosure and show the important role of the motive implied by disclosure in audience evaluations of AI use. Third, we identify perceived legitimacy as a key psychological mechanism for reduced trust and AI acceptance. Finally, this study sheds light on how people's predispositions toward AI moderate these effects, offering a more nuanced understanding of audience reception.

Our findings also hold several implications for journalistic practice. First, even when timely scene photos are unavailable and generative AI presents a convenient solution, newsrooms should be cautious in using AI-generated visuals in disaster coverage. The declining trust and AI acceptance suggest that the reputational cost may outweigh what transparency is supposed to strengthen. Second, if AI-generated images are used, newsrooms should move beyond generic "AI-generated" labels and instead frame AI disclosures in ways that resonate with public interest and journalistic ethics, framing AI use with altruistic justifications to minimize the reputational cost of AI disclosures. Furthermore, our results highlight the critical role of perceived legitimacy in shaping audience attitudes toward AI-assisted journalism. Effective AI disclosures should not only inform audiences of AI involvement but also foster perceptions that AI usage is not opportunity-driven but ethically sound and professionally justified. Last,

AI disclosure strategies should be tailored to diverse publics with different predispositions toward AI. For AI enthusiasts, framing AI use from an altruistic perspective may enhance trust and AI acceptance; for AI skeptics, newsrooms might avoid using AI-generated visuals. Developing tailored disclosure strategies could help foster trust and AI acceptance in the future of AI-assisted journalism.

Limitations

This study is not without limitations. First, our investigation focuses on disaster news coverage, a high-stakes context that might heighten public expectations for journalistic ethics and trigger harsher judgments than in low-stakes topics, such as entertainment or sports. Future studies should examine AI disclosure effects across diverse issue contexts to determine whether the observed audience backlash is unique to crisis reporting.

Second, drawing from the identifiable victim effect, we operationalized victim identifiability as victim count (single vs. multiple), which did not significantly affect audience responses, but it does not rule out the influence of victim representation in AI-generated images. Alternative visual cues, such as content type (photorealistic or animated) or the presence of human faces, may more effectively shift people's moral judgments of AI use and warrant future investigations.

Third, our study operationalized AI disclosure in a relatively narrow way by focusing on attributed motives for AI use. This limits the external validity of the results to contexts where disclosures resemble simple labeling or brief justifications, rather than more ecologically valid, human-in-the-loop formulations. For example, Altay and Gilardi (2024) found that disclosures emphasizing human journalists' supervision in the production process can reduce some negative reactions to AI-generated news headlines, compared with disclosures suggesting that AI fully replaces human journalists. Future research should examine whether other forms of AI disclosure, such as highlighting human supervision or explaining how AI was used, shape public responses differently. Additionally, placing disclosure messages in image captions may not be the most effective approach for conveying source information and may be neglected by some audiences. Future work should test alternative presentation strategies, such as adding the editor's notes or prominent labels.

Fourth, although Prolific generally provides higher-quality online samples than other crowdsourced platforms (Peer et al., 2021), its samples tend to skew female, younger, and more educated than the general population (Prolific, 2025). Our sample similarly showed this biased pattern. Since such groups may differ from older or less educated populations in AI literacy and attitudes toward AI, future studies should recruit more representative samples to better capture audience responses to AI-generated news images.

Fifth, although we assessed attitudes toward AI prior to stimulus exposure to avoid post-treatment bias, this premeasurement may have heightened participants' sensitivity to the AI-generated elements in the stimuli, thereby limiting the causal interpretation of the moderation results. Future research should measure attitudes toward AI at an earlier time point, with a sufficient temporal gap before exposure, to minimize

potential priming effects and ensure that the pre-measure does not influence how participants process the stimuli.

Finally, our study treats AI disclosure as a default practice, yet journalists can choose not to disclose AI usage in practice. Prior research suggests that third-party revelations of undisclosed AI usage can reduce trust even more than self-disclosure (Schilke & Reimann, 2025). Future research should identify the specific contexts in which the cost of potential third-party exposure may outweigh the reputational damage resulting from self-disclosure.

ORCID iDs

Yuhan Li https://orcid.org/0000-0002-8992-1601
Hang Lu https://orcid.org/0000-0002-4795-3410

Ethical Considerations

This study was approved by the Institutional Review Boards of University of Michigan, and all procedures complied with relevant ethical guidelines.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Chair's Graduate Student Research Grant and the Science and Environmental Communication Research Group, Science and Communication Research Funding (Grant No. Go31843), both at the University of Michigan—Ann Arbor.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data underlying this article will be shared on reasonable request to the corresponding authors.

Supplemental Material

Supplemental material for this article is available online.

References

Altay, S., & Gilardi, F. (2024). People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation. *PNAS Nexus*, *3*(10), pgae403. https://doi.org/10.1093/pnasnexus/pgae403

Baek, T. H., Kim, J., & Kim, J. H. (2026). Effect of disclosing AI-generated content on prosocial advertising evaluation. *International Journal of Advertising*, *45*(1), 171–192. https://doi.org/10.1080/02650487.2024.2401319

Bewersdorff, A., Hornberger, M., Nerdel, C., & Schiff, D. S. (2025). AI advocates and cautious critics: How AI attitudes, AI interest, use of AI, and AI literacy build university students'

AI self-efficacy. *Computers and Education: Artificial Intelligence*, *8*, 100340. https://doi.org/10.1016/j.caeai.2024.100340

Blassnig, S., Mitova, E., Strikovic, E., Urman, A., de Vreese, C., Hannák, A., & Esser, F. (2024). User perceptions of news recommender systems and trust in media outlets: A five-country study. *Journalism Studies*, *25*(10), 1182–1204.

Carlson, M. (2017). *Journalistic authority: Legitimating news in the digital era*. Columbia University Press.

Coombs, W. T. (2007). Attribution theory as a guide for post-crisis communication research. *Public Relations Review*, *33*(2), 135–139. https://doi.org/10.1016/j.pubrev.2006.11.016

Culver, K. B., & Lee, B. (2019). Perceived ethical performance of news media: Regaining public trust and encouraging news participation. *Journal of Media Ethics*, *34*(2), 87–101. https://doi.org/10.1080/23736992.2019.1599720

Dan, V., Paris, B., Donovan, J., Hameleers, M., Roozenbeek, J., Van Der Linden, S., & Von Sikorski, C. (2021). Visual mis-and disinformation, social media, and democracy. *Journalism & Mass Communication Quarterly*, *98*(3), 641–664. https://doi.org/10.1177/10776990211035395

de Fine Licht, J., Naurin, D., Esaiasson, P., & Gilljam, M. (2014). When does transparency generate legitimacy? Experimenting on a context-bound relationship. *Governance*, *27*(1), 111–134. https://doi.org/10.1111/gove.12021

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. *AI & Society*, *35*(4), 917–926. https://doi.org/10.1007/s00146-020-00960-w

Deephouse, D. L., & Suchman, M. (2008). Legitimacy in organizational institutionalism. In R. Greenwood, C. Oliver, R. Suddaby, & K. Sahlin (Eds.), *The SAGE handbook of organizational institutionalism* (pp. 49–77). SAGE Publications Ltd. https://doi.org/10.4135/9781849200387.n2

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, *5*(7), 809–828. https://doi.org/10.1080/21670811.2016.1208053

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Grassini, S. (2023). Development and validation of the AI attitude scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. *Frontiers in Psychology*, *14*, 1191628. https://doi.org/10.3389/fpsyg.2023.1191628

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.

Heider, F. (1958). *The psychology of interpersonal relations*. John Wiley & Sons. https://doi.org/10.1037/10628-000

Helberger, N., van Drunen, M., Moeller, J., Vrijenhoek, S., & Eskens, S. (2022). Towards a normative perspective on journalistic AI: Embracing the messy reality of normative ideals. *Digital Journalism*, *10*(10), 1605–1626. https://doi.org/10.1080/21670811.2022.2152195

Iyengar, S. (1990). Framing responsibility for political issues: The case of poverty. *Political Behavior*, *12*(1), 19–40. https://doi.org/10.1007/BF00992330

Jahn, J., Eichhorn, M., & Brühl, R. (2020). How do individuals judge organizational legitimacy? Effects of attributed motives and credibility on organizational legitimacy. *Business & Society*, *59*(3), 545–576. https://doi.org/10.1177/0007650317717959

Jeong, S.-H. (2009). Public's Responses to an oil spill accident: A test of the attribution theory and situational crisis communication theory. *Public Relations Review*, *35*(3), 307–309. https://doi.org/10.1016/j.pubrev.2009.03.010

Ji, Y., Tao, W., & Wan, C. (2025). A systematic review of attribution theory applied to crisis events in communication journals: Integration and advancing insights. *Communication Research*. Advance online publication. https://doi.org/10.1177/00936502251319843

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). Academic Press. https://doi.org/10.1016/S0065-2601(08)60107-0

Karlsson, M. (2010). Rituals of transparency: Evaluating online news outlets' uses of transparency rituals in the United States, United Kingdom and Sweden. *Journalism Studies*, *11*(4), 535–545. https://doi.org/10.1080/14616701003638400

Kelly, S., Kaye, S.-A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, *77*, 101925. https://doi.org/10.1016/j.tele.2022.101925

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lu, H. (2024). Highlighting victim vividness and external attribution to influence policy support regarding the opioid epidemic: The mediating role of emotions. *Health Communication*, *39*(7), 1333–1342. https://doi.org/10.1080/10410236.2023.2212139

Martin, K., & Waldman, A. (2023). Are algorithmic decisions legitimate? The effect of process and outcomes on perceptions of legitimacy of AI decisions. *Journal of Business Ethics*, *183*(3), 653–670. https://doi.org/10.1007/s10551-021-05032-7

Martinez, L., Caramiaux, B., & Fdili Alaoui, S. (2025). Generative AI in documentary photography: Exploring opportunities and challenges for visual storytelling. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). *Association for Computing Machinery*. https://doi.org/10.1145/3706598.3714200

Miotto, G., & Youn, S. (2020). The impact of fast fashion retailers' sustainable collections on corporate legitimacy: Examining the mediating role of altruistic attributions. *Journal of Consumer Behaviour*, *19*(6), 618–631. https://doi.org/10.1002/cb.1852

Mitova, E., Blassnig, S., Strikovic, E., Urman, A., Hannak, A., de Vreese, C., & Esser, F. (2025). Exploring public attitudes toward generative AI for news across four countries. *Journal of Quantitative Description: Digital Media*, *5*, Article 012. https://doi.org/10.51685/jqd.2025.012

Morosoli, S., Resendez, V., Naudts, L., Helberger, N., & de Vreese, C. (2025). "I resist": A study of individual attitudes towards generative AI in journalism and acts of resistance, risk perceptions, trust and credibility. *Digital Journalism*, *14*(5), 771–790. https://doi.org/10.1080/21670811.2024.2435579

New York Times. (2024, October 7). How the New York Times uses A.I. for Journalism. https://www.nytimes.com/2024/10/07/reader-center/how-new-york-times-uses-ai-journalism.html

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Prolific. (2025). *What are the advantages and limitations of an online sample?* Prolific.com; Prolific Research. https://researcher-help.prolific.com/en/articles/445274-what-are-the-advantages-and-limitations-of-an-online-sample

Ramasubramanian, S. (2011). The impact of stereotypical versus counterstereotypical media exemplars on racial attitudes, causal attributions, and support for affirmative action. *Communication Research*, *38*(4), 497–516. https://doi.org/10.1177/0093650210384854

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, *188*, 104405. https://doi.org/10.1016/j.obhdp.2025.104405

Slovic, P. (2007). "If I look at the mass, I will never act": Psychic numbing and genocide. *Judgment and Decision making*, *2*(2), 79–95.

Strikovic, E., & Cools, H. (2025). Reality re-imag(in)ed. mapping publics' perceptions and evaluations of AI-generated images in news contexts. *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2025.2573076

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, *20*(3), 571–610. https://doi.org/10.2307/258788

Thomson, T. J., Thomas, Ryan J., & Matich, P. (2025). Generative visual AI in news organizations: challenges, opportunities, perceptions, and policies. *Digital Journalism*, *13*(10), 1693–1714. https://doi.org/10.1080/21670811.2024.2331769

Toff, B., & Simon, F. M. (2024). "Or they could just not use it?": The dilemma of AI disclosure for audience trust in news. *The International Journal of Press/Politics*, *30*(4), 881–903. https://doi.org/10.1177/19401612241308697

Tong, J. (2018). Journalistic legitimacy revisited: Collapse or revival in the digital age? *Digital Journalism*, *6*(2), 256–273. https://doi.org/10.1080/21670811.2017.1360785

von Sikorski, C., & Hameleers, M. (2025). Disinformation in the age of artificial intelligence (AI): Implications for journalism and mass communication. *Journalism & Mass Communication Quarterly*, *102*(4), 941–957. https://doi.org/10.1177/10776990251375097

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*(9), 1115–1125. https://doi.org/10.1177/01461672992512005

Weiner, B. (1986). *An attributional theory of motivation and emotion*. Springer US. https://doi.org/10.1007/978-1-4612-4948-1

Weiner, B. (2006). *Social motivation, justice, and the moral emotions: An attributional approach*. Psychology Press. https://doi.org/10.4324/9781410615749

Wittenberg, C., Epstein, Z., Péloquin-Skulski, G., Berinsky, A. J., & Rand, D. G. (2025). Labeling AI-generated media online. *PNAS Nexus*, *4*(6), pgaf170. https://doi.org/10.1093/pnasnexus/pgaf170

Zelizer, B. (2010). *About to die: How news images move the public*. Oxford University Press.

Author Biographies

Yuhan Li (M.A., Tsinghua University) is a Ph.D. candidate in the Department of Communication and Media at the University of Michigan—Ann Arbor. She is interested in science communication, environmental communication, and computational social science. She primarily uses experimental and computational methods to study public understanding and acceptance of science (e.g., climate change) and technologies (e.g., AI) and how people's understanding of science and technology drives their perceptions, attitudes, and behaviors.

Hang Lu (Ph.D., Cornell University) is an Associate Professor in the Department of Communication and Media, University of Michigan—Ann Arbor. His research revolves around understanding how different audience segments respond to media messages regarding science, health, environmental, and risk issues, and how these messages can be crafted to maximize their effectiveness.