International Journal of Communication 19(2025), 3691–3711 1932–8036/20250005

A Beacon of Trustworthiness in a Sea of Disinformation: Does News Coverage About the Dangers of Generative AI Cause People to Flock to Journalism?

TOM DOBBER
MICHAEL HAMELEERS
CHRISTOPHER STARKE
TONI VAN DER MEER
University of Amsterdam, The Netherlands

Despite its merits in advancing and simplifying information flows, generative AI is often framed in the news as a formidable tool for advancing disinformation. Through the lens of truth-default theory, this study argues that negatively framed coverage about generative AI makes people more uncertain about what is true. Subsequently, people attempt to resolve that uncertainty by relying on verified information: journalism. This study investigates how differentially framed news articles about the disinformation potential of generative AI can cause people to flock to journalism. Contrary to our expectations, an online experiment (N = 658) among Dutch participants indicates that those exposed to a negatively framed (alarmist and loss) article about generative AI do not become uncertain but do become less trusting toward journalism overall. The results suggest that people do not perceive journalism as a trustworthy solution to the potential disinformation problems caused by generative AI. Moreover, our findings suggest that emphasizing the risks of AI-driven deception affects how people perceive the information environment as a whole, rather than explicitly distinguishing between journalism and nonjournalistic content.

Keywords: generative AI, trust in journalism, truth-default theory, framing

Journalistic trust is globally under pressure (Newman, Fletcher, Eddy, Robertson, & Nielsen, 2023), threatening a well-informed citizenry and well-functioning democracy (Downs, 1957; Habermas, 2005). Disinformation is a key factor potentially contributing to the erosion of trust in journalism (Vaccari & Chadwick, 2020; Zimmermann & Kohring, 2020). Especially in light of emerging technologies, such as (generative) artificial intelligence (AI), disinformation can be presented in a realistic manner, as exemplified

Tom Dobber: t.dobber@uva.nl
Michael Hameleers: m.hameleers@uva.nl
Christopher Starke: c.d.r.o.starke@uva.nl
Toni van der Meer: g.l.a.vandermeer@uva.nl
Date submitted: 2025-02-20

in the case of deepfakes. While the literature does not agree on generative AI's inevitable negative contribution to our information ecosystem (Simchon, Edwards, & Lewandowsky, 2024; Simon, Altay, & Mercier, 2023), news media often highlight AI threats using negative frames (Fried, 2023; Milmo & Hern, 2023). In this study, we bring forward and test the notion that news coverage about the dangers of generative AI to our information ecosystem has a positive spillover effect on people's generalized trust in journalism. Specifically, we argue that negatively framed news coverage makes people uncertain about what is true and what is not outside the realm of established news, and as a coping mechanism, people turn to verified information (i.e., journalism) to resolve that uncertainty.

Research suggests that disinformation is sparse in the average person's information diet, with estimates ranging between 1% and 6% in the United States and Europe (Acerbi, Altay, & Mercier, 2022). In line with these optimistic levels of deception, truth-default theory posits that most communication is honest and truthful, and that people are more likely to rate novel information as honest than dishonest (Levine, 2014). This truth default is beneficial because, although people can occasionally be fooled, communication is more efficient when we do not need to critically evaluate all encountered information.

Yet, people abandon the truth-default state when suspicion is triggered. Levine argues that "information from a third-party warning of potential deception" can act as a trigger (Levine, 2014, p. 386). Triggers are prominent today, as the media and other established institutions ring the alarm on deception and disinformation, especially connected to emerging technologies such as AI. The problem is that the dominance of these negative frames may not correspond to the actual small salience of the problem in people's general media diets, potentially causing inflated risk perceptions and perceived disinformation prevalence.

Recent research on disinformation suggests that informing people about potential disinformation can make them more skeptical of all incoming information (true and false; Ternovski, Kalla, & Aronow, 2022; Van der Meer, Hameleers, & Ohme, 2023). In other words, disinformation warnings might cause people to abandon the truth-default state. Considering that exposure to disinformation is relatively rare (Acerbi et al., 2022) and that warnings of disinformation might cause people to overestimate its prevalence, it becomes apparent that disinformation is, in part, a *perceptual* problem. The media may play a central role in fueling these perceptions, specifically through framing (Entman, 1993). Especially when reporting on emerging issues unfamiliar to people, the news media can shape how people think about said issues, as people have limited knowledge and cognition to process news frames (Han, Chock, & Shoemaker, 2009).

In this preregistered experimental study,[1] we focus on an emerging issue surrounded by negative frames in the public discourse: generative AI. Generative AI allows people to generate both true and false text, images, and audio based on text prompts.

This experimental study (N = 658 in the Netherlands) pursues three objectives. First, we explore whether exposure to media coverage about generative AI triggers people to abandon their truth-default state. Second, we investigate the consequences of exiting the truth-default state by examining the degree

[1] https://osf.io/7qkdh/?view_only=522aa6eb9a46486d959cca6986aab434

to which people become uncertain of what is true and what is false. Third, we aim to understand how people respond to their hypothesized state of uncertainty. Whereas truth-default theory states that people who abandon their truth-default state will not retain their new state indefinitely (Levine, 2014), this study tests the idea that uncertain people will attempt to resolve or reduce their uncertainty by flocking toward verified journalistic information.

Trigger Events Make People Abandon the Truth-Default State

We define disinformation as goal-directed false information—that is, the creation and dissemination of false information with the intention to deceive recipients (Hameleers, 2023b). In keeping with recent research emphasizing the consequences of disinformation as a label or discourse (Egelhofer, Boyer, Lecheler, & Aaldering, 2022), we specifically focus on how emphasizing the threats of disinformation—related to AI-powered disinformation—affects trust in journalism. Previous studies found that emphasizing the threats of false information can make people more skeptical toward all information (Hoes, Aitken, Zhang, Gackowski, & Wojcieszak, 2024; Van der Meer et al., 2023), and when populist politicians accuse news outlets of being "fake news," citizens' trust in the accused news outlet and related accuracy perceptions drop (Egelhofer et al., 2022).

We build on these studies by postulating that mainstream media coverage emphasizing threats of false information and deception can foster uncertainty about incoming information, subsequently increasing trust in verified information (i.e., journalism). Moreover, we argue that uncertainty can increase people's intention to rely on journalistic information to decrease or resolve it. Thus, while Van der Meer et al. (2023) and Hoes et al. (2024) address how disinformation warnings affect the perceived credibility of factual information, and Egelhofer et al. (2022) focus on disinformation accusations targeting a specific news outlet and article, this study takes a step back and focuses on journalism as a whole, more indirectly via uncertainty. It does not mention a specific news outlet or warn of specific information. Instead, it highlights the threats to our information ecosystem posed by the rise of generative AI. In addition, we apply these insights to negatively framed (i.e., alarmistic and loss) communication about an allegedly high-risk development in information dissemination: generative AI. We explain the potential effects of such negative frames on disinformation risks through the lens of truth-default theory (Levine, 2014).

The truth-default state holds that people have a passive bias toward the truth. By default, people generally accept incoming information as truthful without questioning its veracity (Levine, 2014). However, certain triggers can pull people of the truth-default state, enabling them to detect deception.

Trigger events include, but are not limited to (a) a projected motive for deception, (b) behavioral displays associated with dishonest demeanor, (c) a lack of coherence in message content, (d) a lack of correspondence between communication content and some knowledge of reality, or (e) information from a third-party warning of potential deception. (Levine, 2014, p. 386)

In this study, we focus specifically on trigger E because this trigger is most likely encountered through news consumption about generative AI. Triggers A to D, by contrast, rely more strongly on individuals' own assessments than on those of third parties, such as news media.

In line with this trigger, disinformation literature suggests that warning people about disinformation through media literacy interventions might backfire. Such interventions might create the perception that disinformation is omnipresent, leading to skepticism toward all information, both true and false (Hoes et al., 2024; Ternovski et al., 2022; Van der Meer et al., 2023). Thus, by sounding the alarm on mis- and disinformation, and by pointing people to seemingly high degrees of deception in their information environment, news about disinformation may motivate people to deviate from the truth default by acting as a trigger.

In line with this premise, Ternovski et al. (2022) conducted an experiment using a simple warning message about deepfakes with a credibility cue and found that warned participants started to distrust both real and manipulated videos. Van der Meer et al. (2023) found that people exposed to a general warning about the presence of misinformation also increased skepticism toward factual news. Hameleers (2023a) used prebunking (i.e., a preemptive warning) media literacy interventions and found that participants exposed to such messages were more skeptical of factually accurate information than participants who were not. Hoes et al. (2024) found that exposure to media literacy interventions and news coverage of misinformation increased overall skepticism. Similarly, Egelhofer et al. (2022) found that fake news labels and disinformation accusations lowered trust in the accused news outlet, underscoring how triggering deception by warning about false content can have unintended effects beyond lowering the credibility of disinformation itself.

Together, extant literature suggests that discussing the threats of disinformation or using labels of deception may cause distrust in information and heighten the perceived threats of deceptive information. This evidence on deception primes will be considered in the hypotheses.

Media Literacy Interventions May Inadvertently Trigger Suspicion

Negative effects of warning about false information found in previous studies could partially be driven by the (relatively straightforward) type of warning message used. Previous studies on the side effects of mis- and disinformation warnings are mostly based on the idea that only stating the dangers of false information would trigger suspicion, without considering a more comprehensive "inoculation" approach in which a warning message is paired with exposure to a small dose of deception (i.e., the threat that is warned against).

In a meta-analysis on the adverse effects of inoculation against misinformation (Lu, Hu, Li, Bi, & Ju, 2023), a contrary point seems to arise: inoculation interventions appear not to spillover to factually accurate information. The meta-analysis found that inoculation interventions affected people's "Real Information Credibility Assessment" (d = 0.20). Additionally, the meta-analysis determined that "Credibility Discernment" also improved significantly (d = 0.20), highlighting the potential of successfully inoculating people against misinformation.

However, the existing evidence is too limited to conclude that inoculation does not increase skepticism toward factually accurate information in general. The meta-analysis (Lu et al., 2023) included studies that inoculated people about a specific claim or a specific piece of disinformation and subsequently

checked the degree to which people found accurate claims about that specific topic or actor trustworthy. While a valuable scientific contribution, it is too narrow to decisively refute the argument that general media literacy interventions, including those using inoculation, make people skeptical of all information, both true and false. Thus, we argue that any media literacy intervention (including inoculation) might function as a trigger by inadvertently creating the perception that disinformation is more prevalent than empirical research suggests (Acerbi et al., 2022). Although we do not use an inoculation intervention in this study, as such interventions pair exposure to disinformation with a warning message and suggestions for detecting deception, we do focus on one central aspect of inoculation interventions: warning people about the threats of deceptive information and exposing them to content that suggests deception is prevalent and requires a critical outlook on incoming information.

News Articles as a Trigger: Framing Threats in Media Coverage

Efforts to increase literacy among citizens rely on interventions such as games, texts, images, videos, or more classical educational setups. This study argues that news media coverage about generative AI can also serve as a de facto media literacy intervention. People learn about emerging technologies through the media (Cacciatore et al., 2012). The media use emphasis frames that suggest a certain problem interpretation (Entman, 1993) and can influence how people think about issues (Tversky & Kahneman, 1986; Valkenburg, Semetko, & de Vreese, 1999), such as generative AI. By emphasizing negative elements, such as the potential to generate realistic disinformation, news coverage can affect people's knowledge of and resilience toward disinformation, similar to a classic media literacy intervention. As such, we argue that negatively framed (i.e., alarmistic and loss-framed) news articles about generative AI may trigger people to abandon the truth-default state. In doing so, negatively framed news articles act as a de facto media literacy intervention. News articles about generative AI that are not framed negatively do not qualify as triggers because they do not warn people about potential deception.

In this study, we benchmark a neutral frame (factual description, not emphasizing threats or opportunities), a gain frame (emphasizing the potential to gain something in relation to generative AI), a loss frame (emphasizing the potential to lose something related to generative AI), an alarmist frame (sounding the alarm about generative AI, more extreme condition than loss frame), and a relativizing frame (emphasizing that generative AI offers threats and opportunities, and that the threats are not so threatening) against an unrelated neutral news article that serves as the control condition. Crucially, we expect news articles containing a loss or alarmist frame to serve as a trigger, as they include a general warning about generative AI's potential to deceive (in line with Van der Meer et al., 2023). Whereas these two negative frames serve as warning messages, the remaining frames (i.e., gain, relativizing, neutral) illustrate news coverage that is not expected to function as a trigger. In other words, only the negative frames are expected to trigger people of the truth-default state. However, what happens when people exit the truth-default state remains poorly understood.

Truth Default Versus Uncertainty Default

According to truth-default theory, leaving the truth-default state is temporary; once a person enters a state of suspicion, they eventually pivot back to the truth default (Levine, 2014). Hameleers

(2023a) offers an alternative scenario and suggests that "the (potentially disproportionate) attention to mis- and disinformation" (p. 4) may shift people from the truth-default state to a deception-default state when facing a veracity judgment.

However, different from Hameleers (2023a), this study argues that exposure to a deception trigger propels people into a state of uncertainty, as the (limited) empirical evidence might point more strongly to a state of uncertainty than to a deception default. Hameleers (2023a) finds that those exposed to media literacy interventions, on average, scored significantly worse on perceived accuracy of true information than those who did not. Hameleers (2023a) concluded that "discussions about mis- and disinformation may be an important trigger event. Hence, (..) public and politicized debates on how to 'fight' mis- and disinformation may contribute to a gradual shift toward a deception bias" (p. 8). However, while the experimental participants in Hameleers (2023a) scored significantly worse than the control condition, the experimental participants still scored close to the midpoint of the information credibility scale (M = 3.95; SD = .98 on a 7-point information credibility scale), suggesting that people moved toward a state of uncertainty rather than into a state of deception default. Moreover, Vaccari and Chadwick (2020) demonstrated that deceptive communication can sow uncertainty. Exposure to a deepfake did not necessarily deceive people; rather, it increased uncertainty about the information presented in the deepfake (Vaccari & Chadwick, 2020). Finally, Newman, Fletcher, Robertson, Ross Arguedas, and Nielsen (2024) concluded that people are generally uncertain about differentiating between true and false information online, a condition that reflects a state of uncertainty rather than deception.

Uncertainty can occur when citizens perceive communication toward them as deficient (Alvarez & Brehm, 1997). For example, when they learn that bad faith actors can create realistic pieces of disinformation at scale through generative AI. In accordance with truth-default theory (Levine, 2014), learning from a third party about a new tool for deceit may trigger people to abandon their truth-default state. We argue that this shift places them in a state of uncertainty.

Journalistic Principles

Generally, uncertainty is a state people want to remove or at least reduce (Downs, 1957; Hogg, 2000) because uncertainty is an uncomfortable state that is "associated with reduced control over one's life" (Hogg, 2000, p. 227). Reducing uncertainty can be achieved "through the acquisition of information" (Alvarez & Brehm, 1997; Downs, 1957, p. 77). Of course, when one learns that realistic disinformation can be generated on a large scale, information acquisition becomes risky. Indeed, Vaccari and Chadwick (2020) found that exposure to a deepfake made people not only more uncertain but also less trusting of "news and information about politics and public affairs that you see on social media" (p. 5). People do not trust news on social media because anyone can post information, and it is often uncertain whether such information is verified (Newman & Fletcher, 2017).

This is where journalism stands out compared with social media news: verifying information is one of the core journalistic principles that citizens cite as a reason for trusting journalism (Newman & Fletcher, 2017). Journalistic principles set journalism apart from nonjournalistic news by increasing the rigor of

reporting and ensuring the reliability of news coverage. Alongside verification, one can think of independence, transparency, and objectivity as its key principles.

This study argues that a negatively framed article about the disinformation-related dangers of generative AI functions as a deception trigger, causing people to abandon the truth-default state (Levine, 2014), become uncertain about what is real, and consequently find it more difficult to discern between untrustworthy and honest information. To resolve that uncertainty, people are expected to turn to journalism as an alternative, established information format driven by objectivity, balance, and neutrality in perspectives. Given that established journalism is known for its principles that include verifying information (Newman & Fletcher, 2017), people who rely on journalistic sources need to be less uncertain about whether the information is AI generated or not. Indeed, in crisis communication, threatening situations cause high information need and increase uncertainty about the threatening situation. To fulfill the information need and reduce uncertainty, people turn to information sources that rely on journalistic standards (Van der Meer, 2018). Thus, we propose that the triggered uncertainty can be resolved or reduced by approaching journalism.

Journalistic Knowledge

We argue that this effect of uncertainty on journalistic trust and the intention to rely on journalistic sources is stronger for people with greater knowledge of journalism: These people are more likely to be familiar with journalistic principles designed to increase information quality and are thus better able to distinguish between journalistic and nonjournalistic information. We conceptualize these journalistic principles as transparency, verification of information, objectivity, and independence. Studies show that people tend to trust journalistic news media because of (one of) these principles (Newman & Fletcher, 2017), conceptualize trust in journalism along the lines of those values (Knudsen, Dahlberg, Iversen, Johannesson, & Nygaard, 2022), or mention these principles as indicators of "good journalism" (Gil de Zúñiga & Hinsley, 2012). People who are more knowledgeable about journalism are assumed to be more aware of the existence and usefulness of these journalistic values. In other words, people who, for instance, do not know that journalism strives to be objective and independent have more trouble understanding why journalistic news is often more trustworthy than news that does not apply journalistic principles. As such, if they were to experience uncertainty, they would not know that journalistic news is reliable information, which means that it is unlikely that they will resolve their uncertainty by "flocking to journalism." As such, we propose the following model (see Figure 1). The hypotheses are listed after Figure 1.

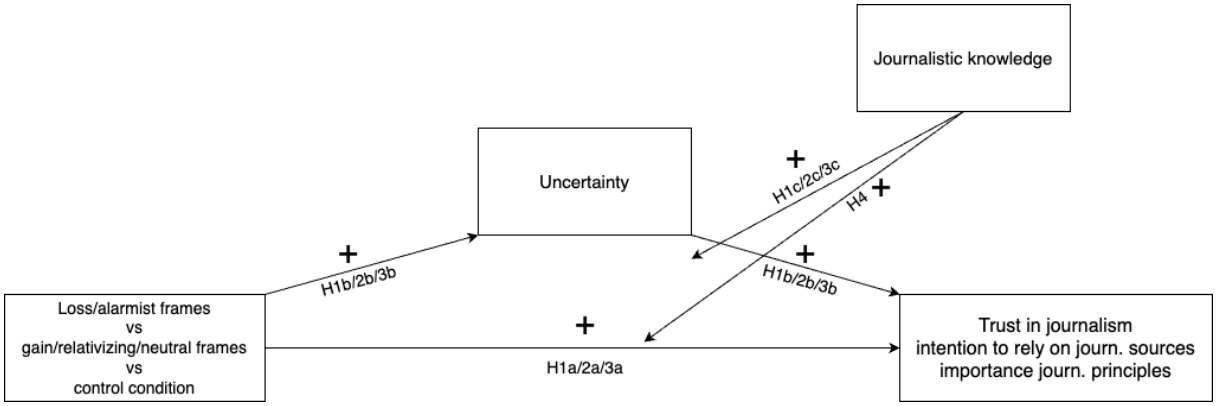

Figure 1. Schematic drawing of the expected relationships.

Hypotheses Relating to Trust

H1a. *Exposure to negatively framed news (loss frame and alarmist frame vs. gain, relativizing, neutral frame, and control condition) about generative AI increases trust in journalism.*

H1b. *This effect is mediated by uncertainty, so that more uncertainty leads to more trust in journalism.*

H1c. *The mediated effect of uncertainty on trust in journalism will be moderated by journalistic knowledge, so that the effect is stronger for people with more journalistic knowledge.*

Hypotheses Relating to Intention to Rely on Journalistic Sources

H2a. *Exposure to negatively framed news (loss frame and alarmist frame vs. gain, relativizing, neutral frame, and control condition) about generative AI increases intention to rely on journalistic sources.*

H2b *This effect is mediated by uncertainty, so that more uncertainty leads to a higher intention to rely on journalistic sources.*

H2c. *The mediated effect of uncertainty on intention to rely on journalistic sources will be moderated by journalistic knowledge, so that the effect is stronger for people with more journalistic knowledge.*

Hypotheses Relating to Intention to Journalistic Principles

H3a. *Exposure to negatively framed news (loss frame and alarmist frame vs gain, relativizing, neutral frame, and control condition) about generative AI increases perceived importance of journalistic principles.*

H3b. *This effect is mediated by uncertainty, so that more uncertainty leads to higher perceptions of the importance of journalistic principles.*

H3c. *The mediated effect of uncertainty on perceived importance of journalistic principles will be moderated by journalistic knowledge, so that the effect is stronger for people with more journalistic knowledge.*

Hypothesis Relating Knowledge as a Moderator and Trust as an Outcome

H4. Exposure to negatively framed news (loss frame and alarmist frame) about generative AI increases trust in journalism, and this effect is stronger when people have more knowledge of journalism.

Method

We conducted a preregistered online experiment (see Appendix F for deviations from the preregistration[2]). After approval from the University of Amsterdam ethical committee, survey company Dynata recruited 706 participants in the Netherlands between June 6 and June 9, 2023. Participants received €2.50, which translates to €11.53 per hour. This is in line with the legal minimum hourly wage of €11.16. We removed speeders (people who completed the survey in less than 180 seconds; n = 43) and flatliners (people with no variation in their answers; n = 5), leaving us with n = 658 participants who finished the survey, on average, in 13 minutes (SD = 35 minutes). There are more than 106 participants per condition (an a priori power analysis run with G*Power showed that each condition required 81 participants to detect an effect size of 0.10, since small effect sizes are commonly found in misinformation research; linear multiple regression, fixed model, single regression coefficient). On average, the sample was 47.7 years old (SD = 16.30), and 51.5% of the sample identified as women. About education, 18% of the sample held a low education level (n = 116), 41% held an intermediate education level (n = 272), and 41% held a high education level (n = 270). This is representative of the Dutch population, which has an average age of 42.4, and consists of 50.3% women. Moreover, in the Netherlands, 26% hold a "low" education level, 35% hold an intermediate level, and 38% hold a "high" education level. Note that this study's sample included only people aged 18 or older. About political orientation, measured on a left-right scale from 0 (left) to 10 (right), the sample is balanced: M = 5.38 (SD = 2.58). Finally, the sample scored an average political interest score of 4.49 (SD = 1.76) on a 7-point scale (7 is the highest).

Design and Stimuli

This experiment consisted of five treatment conditions and one control condition (see Appendix H for a flowchart). Participants were randomly assigned to one of these conditions. Participants in the treatment conditions saw a news article about generative AI. The treatments were framed differently: neutral (n = 108), alarmist (n = 109), relativizing (n = 112), gain frame (n = 107), and loss frame (n = 109). Only the framing differed between the latter four articles; the content of the stimuli remained the same (Appendix A shows the stimuli). The control condition (n = 113) consisted of a neutral news article about elderly workers in the labor market. The stimuli closely resembled an article from the Dutch media brand Nu.nl. Nu.nl was chosen because the brand is well-known and broadly trusted by the Dutch public (Newman, Fletcher, Robertson, Eddy, & Nielsen, 2022). The neutral stimulus counted 127 words, and the stimuli that contained frames ranged between 235 words (alarmistic frame) and 299 words (gain frame). The control condition consisted of 135 words.

[2] https://osf.io/ky5a2/?view_only=288d7227e4db475b99942f8f4f03e06b

Moderators

Knowledge of journalism was measured via the following self-developed true/false knowledge questions: (1) *A journalist is required by law to protect his sources*; (2) *The code of journalism is binding for journalists*; (3) *In the Netherlands, journalist is a protected profession just like lawyer and doctor*; (4) *All journalism in the Netherlands is funded by the Dutch government*; (5) *The subeditor is ultimately responsible for journalistic content and policy*; (6) *Prime Minister Mark Rutte determines what news public broadcaster NOS brings.* All questions should be answered with false. The scale was reversed to enhance interpretability, so that higher scores (6 = highest) reflect greater journalistic knowledge. On average, participants scored 2.09 (SD = 1.27). Four participants scored 6, and 78 participants scored 0.

Mediator

Based on Van der Meer et al. (2023), participants saw four captioned news images and were asked to indicate the extent to which they thought the captioned news image was real (true/false/do not know) and how certain they were of their assessment (0–100 scale; 100 means "absolutely certain"). Unbeknownst to the participants, two captioned news images were real, and two were AI generated. The image was generated using Midjourney, and the caption was generated using ChatGPT. The two real captioned news photos showed a protest of teachers in France and the clearance of an occupied mine in Germany. The AI-generated captioned images depicted a 'unique' picture of US soldiers in Okinawa watching the detonation of the nuclear bomb on Hiroshima, and a close-up of an Extinction Rebellion protestor (see Appendix C). All captioned images were randomly shown to participants. Since we measured uncertainty, we did not check whether people's assessments were correct. After all, if someone gives the wrong answer, but is 100% sure of it, they are not uncertain. However, if someone gives the correct answer, but is 60% sure of it, they are uncertain. The uncertainty scale was calculated by averaging the uncertainty scores across all four captioned images, but only for participants who indicated whether a captioned image was true or false. When participants indicated that they did not know whether a captioned image was true or false, we calculated a 0% certainty score for that specific image. For example, if a person responded to the four images with true (100%), true (80%), false (60%), and do not know (0%), their average uncertainty score would be 60. The average certainty score was 56.43 (SD = 22.87). The two real conditions were rated with an average certainty score of M = 51.02 (SD = 34.25) and M = 63.18 (SD = 31.20), respectively. The two fake conditions scored M = 56.37 (SD = 34.01) and M = 55.14 (SD = 34.13), respectively. As a robustness measure, we also ran our analysis excluding participants who indicated "do not know" on one or more images.

Dependent Variables

Trust in journalism was measured using the Quiring et al. (2021) items, which measured generalized media trust, media skepticism, and media cynicism. The items were altered in two ways for this study. First, the original scale measured news media rather than journalism, so the scale was slightly altered to capture journalism rather than news media. Second, some items were "double-sided" (e.g., "The established news media sometimes are biased, but overall, they reflect the different opinions of the society well"). These were split into two items to ensure clarity. As a result, the scale consists of 14 items in total: seven measuring generalized journalistic trust (Cronbach's alpha = .93; M = 4.85, SD = 1.20), three measuring skepticism

toward journalism (Cronbach's alpha = .76; *M* = 4.66, SD = 1.18), and four items measuring cynicism toward journalism (Cronbach's alpha = .93; *M* = 3.47, SD = 1.69). See Appendix D for the items.

Intended use of journalistic sources was measured using five self-developed items on a 7-point scale. Three items tapped into the intention to consume journalistic work: *Can you indicate the extent to which you are intending to… More often consume news via journalistic sources in the future*, *more often watch NOS Journal in the future*, *more often consume news on social media*. A factor analysis indicated that the last item did not load sufficiently on this dimension, so we removed this item from the scale. Three other 7-point scale items measured people's willingness to pay for journalism: *to what extent are you willing to... pay for access to journalistic news*, *subscribe to a newspaper or journalistic magazine, subscribe to a journalistic news site such as Follow the Money or De Correspondent?* Together, these five items measured intended use of journalistic sources (Cronbach's alpha = .85). Scoring high on this scale signals the intention to increase the use of journalistic sources for news (*M* = 3.56; SD = 1.53).

Importance of journalistic principles was measured using the following four self-developed items on a 7-point scale (7 indicates most important): *I find it important that… news has been verified*; *it is clear how news came to be*; *news is objective*; *news is independent* (Cronbach's alpha = .87; *M* = 5.83; SD = 1.08).

Procedure

Upon entering the survey, participants agreed to the informed consent. 14 did not consent, and their participation was immediately terminated. Then, participants were asked questions about their background (age, gender, education, political interest, left-right political orientation). We then measured participants' political knowledge and asked a single-item question about general trust in news ("In general, I trust the news," 7-point scale, with 7 = highest; *M* = 5.67, SD = 1.58). Participants were randomly exposed to either the stimuli or the control condition (Appendix A). Afterward, we measured their uncertainty scores. Next, participants completed the post survey (question and item order were randomized). Finally, we measured the effectiveness of the manipulation and debriefed participants.

Manipulation Checks

Four items were used to check the extent to which each framed news article came across as intended. Appendix D provides detailed information on the manipulation checks. A series of one-way ANOVAs with Bonferroni corrections showed that conditions were perceived differently—and as intended—on three of the four manipulation check items. Only on the item, *the article was about risks and advantages of AI* did none of the conditions differ significantly, although the control condition had the lowest score on this item. Perhaps this is because of the item's confusing formulation, which focuses on AI's risks and advantages.

Randomization Checks

We conducted randomization checks on age, gender, education, political interest, political orientation (left-right), and the prestimulus measure of general trust in news (trust most media most of the time), and found that randomization for these variables was successful (see Appendix E).

Results

Analytical strategy. First, we analyzed the direct effects of exposure to negatively framed news about generative AI on trust in journalism, intention to rely on journalism, and perceived importance of journalistic principles. Then, we ran PROCESS Model 15 (Hayes, 2022) to examine mediated moderation effects, but only for conditions where ordinary least squares (OLS) regressions reveal a direct effect of X on Y.

The Relationship Between Framing and Trust

A series of OLS regressions using the control group as contrast showed that exposure to specific conditions only affected generalized trust in journalism. Figure 2 shows that participants exposed to an alarmistic frame (B = −.44, t = −2.72, p = .01, 95% CI = [−.75, −.12]) and participants exposed to a loss frame (B = −.35, t = −2.18, p = .03, 95% CI = [−.66, −.03]) scored significantly lower than the control group on their generalized trust perceptions. This contradicts the first part of H1, which expects that exposure to alarmistic or loss-framed news about generative AI would increase rather than decrease trust in journalism. The OLS regressions focusing on skepticism and cynicism yielded no significant differences between the benchmark and experimental conditions (see Figure 3). Thus, our findings suggest that a spillover effect of warning messages occurs rather than enhancing trust in journalism as a coping mechanism, people exposed to negative frames distrust journalism in general.

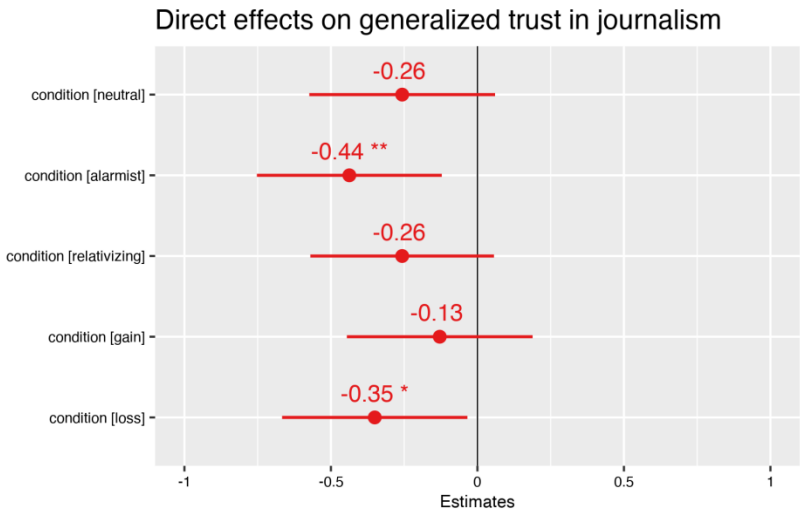

Figure 2. Direct effects of exposure to framed articles on generalized trust in journalism.
Note. OLS regression; control condition is benchmark (vertical line). **p = .01, *p = .03.

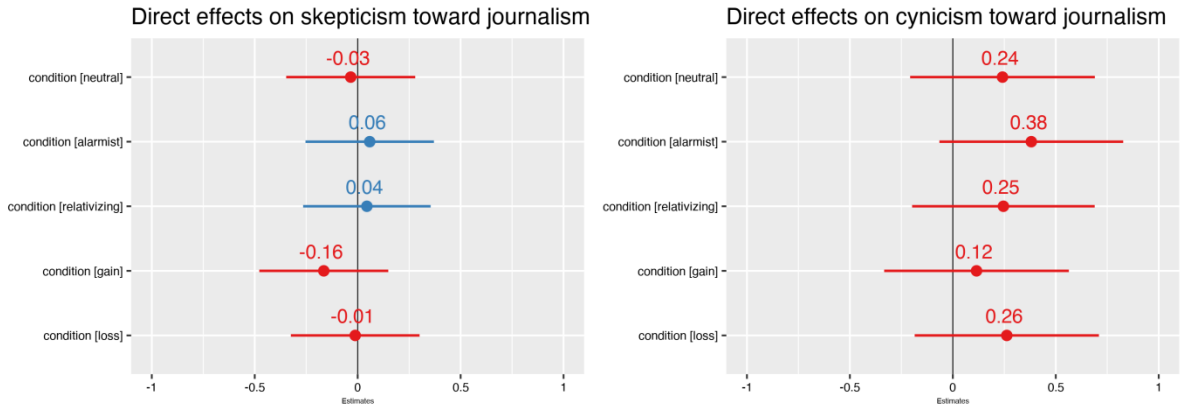

Figure 3. Direct effects of exposure to a framed article on skepticism and cynicism toward journalism.

Note. OLS regression, control condition is benchmark (vertical line). Left panel focuses on skepticism, right panel on cynicism.

Relationship Between Framing and Behavioral Intentions

H2a postulates that exposure to negatively framed news (loss frame and alarmist frame) about generative AI increases participants' intention to rely on journalistic sources. An OLS regression using the control group as contrast showed that no condition differed significantly from the control group (see Figure 4). This means that H2 is not supported: There is no association between any of the news frames and behavioral intentions.

The Relationship Between Framing and the Perceived Importance of Journalistic Principles

In H3a, we expected that people exposed to negatively framed news (loss frame and alarmist frame) about generative AI would perceive journalistic principles as more important compared with the control condition. An OLS regression, however, showed that none of the conditions differed significantly from the control condition (see Figure 4). This means that H3 is not supported.

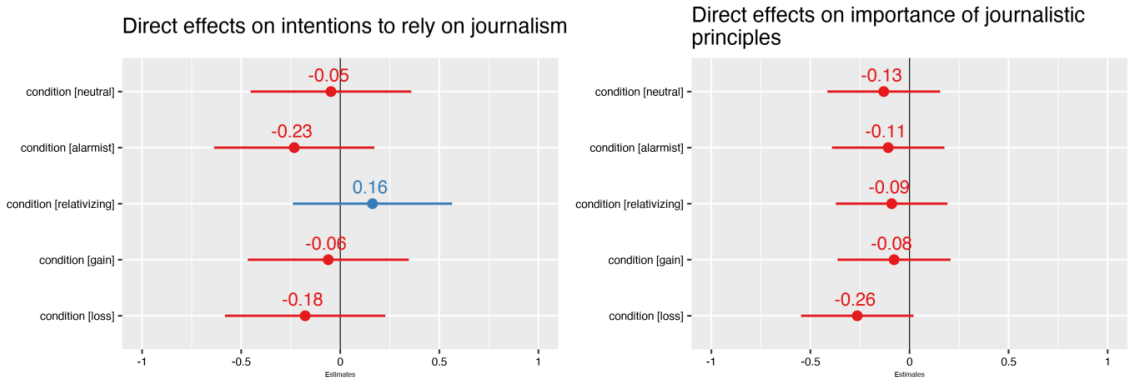

Figure 4. Direct effects of exposure to framed articles on skepticism and cynicism toward journalism.

Note. OLS regression, control condition is benchmark (vertical line). Left panel shows behavioral intentions; right panel shows perceived importance of journalistic principles. Loss condition in the right panel was nonsignificant at $p = .07$.

For all hypotheses, we also ran one-way ANOVAs to examine potential differences between all groups (not purely benchmarked against the control condition), but none of the analyses showed additional significant differences between groups (so between neutral, relativizing, gain, loss, alarmist for the dependent variables generalized journalistic trust, skepticism, cynicism, behavioral intentions, and perceived importance of journalistic principles). In line with the preregistration, we will further study the relationships between exposure to news frames and generalized trust in journalism because we found a direct effect between both variables.

Moderated Mediation Impact of News Frames on Uncertainty and Journalistic Knowledge on Generalized Journalistic Trust

Using PROCESS Model 15 (Hayes, 2022), we estimated two moderated mediation models to compare (1) the alarmistic frame with the control condition, and (2) the loss frame with the control condition. In H1b, we expected that uncertainty would mediate the effect of a negative news frame (alarmistic or loss) on generalized journalistic trust, and in H1c, we expected that the relation between uncertainty and generalized journalistic trust would be moderated by journalistic knowledge (more knowledge means a stronger effect). The analysis of the alarmistic and loss frame conditions yielded no meaningful results. For both conditions, first, uncertainty did not significantly mediate the effect of exposure to negatively framed news on generalized journalistic trust. Second, journalistic knowledge did not significantly moderate the direct effect of exposure to negatively framed news on generalized journalistic trust. As such, the index of moderated mediation was not significant for the alarmistic frame (bootstrapped LLCI: −.04; bootstrapped ULCI: .06) and the loss frame (bootstrapped LLCI: −.02; bootstrapped ULCI: .07). See Appendix G for more information about the models. As such, we conclude that hypotheses H1b, H1c, and H4 are not supported, since exposure to negatively framed articles decreased rather than increased generalized trust in journalism, and since there is no (moderated) mediation of this effect through uncertainty and journalistic knowledge.

The Impact of News Frames on Uncertainty

To ascertain whether exposure to any of the news frames directly affected people's degree of uncertainty, we conducted an OLS regression using the control group as contrast. We found that none of the frames significantly affected people's uncertainty. Figure 5 shows that participants in most experimental conditions (except the loss frame) were more uncertain about the veracity of information, but did not differ significantly from the participants in the control condition. As a robustness check, we excluded the people who answered "do not know" at least once during the measurement of uncertainty (n = 341, leaving a subsample of n = 317). Yet again, the OLS regression on this subsample revealed no significant differences in uncertainty between the individual experimental conditions and the control condition, suggesting robustness. As such, we conclude that exposure to news about generative AI, regardless of framing, does not significantly affect people's uncertainty about the truthfulness of information.

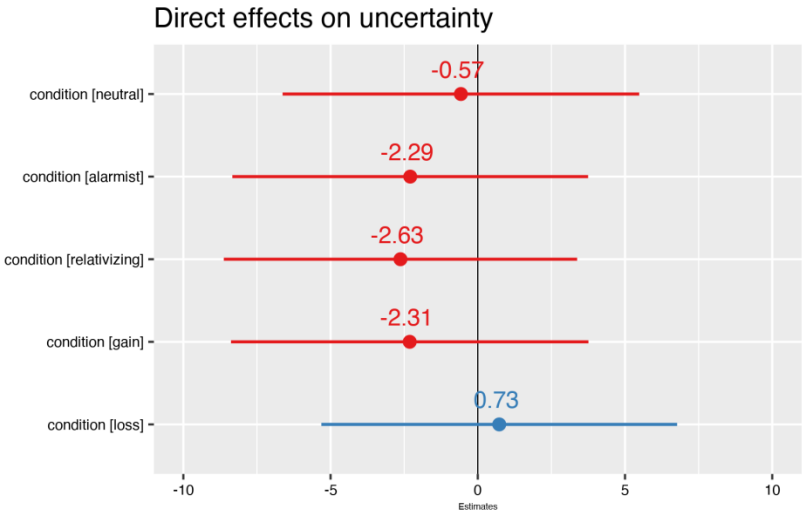

Figure 5. Direct effects of exposure to framed articles on uncertainty.

Note. OLS regression using the control condition as a benchmark (vertical line). Uncertainty was measured on a 0–100 scale; 100 indicated the most certainty.

Discussion

This study set out to understand the consequences of exposure to differently framed media coverage about generative AI. Building upon truth-default theory (Levine, 2014), we sought not only to understand the extent to which such news coverage can trigger uncertainty about what information is real and fake, but also whether it can lead to more trust in and intended reliance on journalism as a coping mechanism to resolve that uncertainty. Contrary to our expectations, we found that exposure to news coverage about generative AI neither significantly increases uncertainty nor significantly increases trust in or reliance on journalism. In fact, exposure to negatively framed news articles (alarmistic frame and loss frame) that stress the disinformation potential of generative AI decreases trust in journalism.

According to truth-default theory (Levine, 2014), people abandon the truth-default state when they encounter a trigger. It is not well understood what happens next, but this study theorized that people become more uncertain about what is real and what is not. However, our analyses show that participants did not become significantly more uncertain after exposure to negatively framed news articles about generative AI. Our findings are more in line with the work of Hameleers (2023a), who suggests that people who exit the truth-default state enter a deception-default state. However, when assuming a deception-default state, one might expect that participants would also score higher on media skepticism and media cynicism, but this is not the case. Perhaps it is more challenging with one intervention to move the needle on skepticism and cynicism than on trust. Moreover, when examining the measurement used in this study, based on Quiring et al. (2021), we observe that the measurement of skepticism and cynicism focuses more on the structural traits of the media. For instance, a skepticism item states: "The established news media sometimes are biased, but overall, they reflect the different opinions of the society well." A cynicism item states, "The established news media and politics conspire to manipulate peoples' opinions." (Quiring et al., 2021, p. 3505). In other words, the operationalization of skepticism and cynicism taps much less directly into doubts about information quality than the operationalization of trust. As such, one could argue that our measurement of skepticism and cynicism was limited, as it focused largely on structural circumstances that were not always directly related to external challenges to the information environment (for this study).

Why Did People Become Less Trusting?

In essence, we expected that when confronted with negatively framed information about the potential of generative AI for disinformation, people would recognize the value of journalistic principles and subsequently turn to a method that generally yields trustworthy information: journalism. The opposite occurred. People did not become uncertain about what was real or not, nor did they display more appreciation for the journalistic method; however, they did become less trusting of journalism.

The most obvious explanation would be that the participants in the negatively framed conditions were still in a suspicion state (Levine, 2014) and did not discriminate between journalism and nonjournalism. Similarly, one could view the news articles as de facto media literacy interventions that taught participants about generative AI and the need to be critical toward information, given that it could be artificial or inauthentic. Media literacy interventions focused on disinformation are often framed in negative terms (e.g., "there is an abundance of problematic information, and this is how you recognize bad information" rather than "there is an abundance of trustworthy information, and this is how you recognize trustworthy information"), and recent research suggests that such media literacy interventions may increase skepticism not only toward false information but also toward true information (Hoes et al., 2024; Van der Meer et al., 2023). Furthermore, the findings from this current study suggest that when people learn about how generative AI might negatively affect our information ecosystem, an alarmistic or loss frame can also negatively spillover toward generalized trust in journalism.

One could also argue that the uncertainty measurement primed participants by showing them how journalists might use generated images inadvertently in their coverage. As a result, people could have become less trusting of journalism. Yet, we are confident that the result shown in Figure 2 is not a pure artifact of our uncertainty measure because in that case, one would expect the relativizing condition and

gain frame conditions to score significantly lower on journalistic trust than the control condition, as both conditions explicitly mentioned the potential for AI to fabricate high-quality images that make it harder to distinguish real from fake. Exposure to the uncertainty measure should then also prime the participants in these two groups, which would drive them to score lower on trust than the control condition (as well as the neutral condition). However, as Figure 2 shows, this is not the case. In addition, the control condition participants (who did not see a framed news article about generative AI) were also exposed to the uncertainty measure. Furthermore, if the uncertainty measure was indeed the driver of the negative effects, one would expect that all conditions, including the control condition, would score similarly statistically, but this was not the case. Indeed, only the negative frame conditions scored significantly different from the control condition. However, while we are certain that the findings are not pure artifacts from the study design, our uncertainty measurement is a limitation of this study because it might have had some priming effect on the participants. Future research could alleviate this concern either by dropping the uncertainty measure altogether or by using content in another setting (e.g., social media).

While the observation of a spillover effect is not without precedence (Hoes et al., 2024; Ternovski et al., 2022; Van der Meer et al., 2023), the findings of this current study beg the question of why reading a negatively framed article did not lead the participants to appreciate journalism as a method that generally yields trustworthy information. It could be that our participants do not agree that journalistic information is generally trustworthy. Indeed, Newman and Fletcher (2017) list several reasons why people do not think news media help them distinguish fact from fiction (e.g., low standards, bias). Swart and Broersma (2023) suggest that youths perceive journalists as necessarily subjective. Possibly, citizens do not distinguish between news that adheres to journalistic standards and news that does not. While youths cognitively understand what news is, this understanding does not align with what they experience as news (Swart & Broersma, 2023). Moreover, Swart and Broersma (2023) found evidence of blurred boundaries between information genres among (some) youths: News could, for instance, also be influencer content or memes if the information is novel. Applying these insights to the current study, we can speculate that participants do not draw a hard line between journalistic information and nonjournalistic information, even though they know the difference. This could explain why the variable journalistic knowledge did not significantly moderate the effect between exposure and generalized journalistic trust: On an affective level, people might struggle to discern journalism from nonjournalism. As such, this study is limited by the absence of a more rounded measure of how people understood 'journalism' and by not controlling for social media use in information seeking.

Losses Versus Gains

Finally, in line with the negativity bias hypothesis, research shows that perceived losses have a larger impact than perceived gains (Fiske, 1980; Tversky & Kahneman, 1986. This could explain why negatively framed articles spillover to trust: people become distrusting of all information because they want to avoid being deceived or misinformed. This study expected that an expected loss or negative outcome would lead to uncertainty, and that uncertainty would be resolved by realizing the strength of journalism. In line with our theoretical argument, the results suggest that people perceive negatively framed articles as warnings. However, different from our theoretical argument, the response does not contain a rational process for resolving uncertainty. It seems that participants respond 'heuristically' by rejecting all

information as a coping mechanism. Consequently, one may wonder when and whether trust levels would "bounce back." After all, truth-default theory stipulates that people will eventually return to the truth-default state (Levine, 2014). This highlights a limitation of this study: we only measured direct effects, so it is possible that we did not capture the entire process. This could be resolved by a repeated-measures experiment, which observes to what extent people either return to their truth-default state or remain in their deception-default state in the medium to long term.

Future research might also explore moderated effects by examining individual-level factors that can predict literacy or vulnerability to AI-generated (mis)information. For example, higher levels of AI literacy or previous exposure to AI-generated content may make people more resilient to threat frames, as they have more skills and confidence in alleviating the threats.

This study was conducted in the Netherlands, which scores relatively high on news trust. News trust has been relatively stable since 2014 (Newman et al., 2024). Approximately 50% to 55% of the Dutch population trusts most news, most of the time. As such, one could argue that in higher-trusting countries, there is more to lose than in lower-trusting countries. This could result in a ceiling effect in lower-trust countries. On the other hand, negative frames about generative AI might further exacerbate or reinforce existing trust perceptions in low-trusting countries. Future research could compare the impact of framing on trust perceptions between countries with higher and lower levels of news trust to test the generalizability of this study's findings across contexts.

Taken together, this study drew upon truth-default theory (Levine, 2014) to examine what happens when people exit the truth-default state by reading negatively framed articles about generative AI. Our findings suggest that negatively framed articles about generative AI undermine generalized trust in journalism itself. By covering the threats to our information environment, news consumers may also associate these threats with journalism, despite journalism's potential to reliably inform readers, even in the face of emerging technologies. As such, journalism might benefit from positioning itself more as an alternative to problematic information caused by AI—as a beacon of trustworthiness in a sea of disinformation and potentially synthetic media.

References

Acerbi, A., Altay, S., & Mercier, H. (2022). Research note: Fighting misinformation or fighting for information? *Harvard Kennedy School Misinformation Review*. doi:10.37016/mr-2020-87

Alvarez, R., & Brehm, J. (1997). Are Americans ambivalent towards racial policies? *American Journal of Political Science, 41*(2), 345–374. doi:10.2307/2111768

Cacciatore, M. A., Anderson, A. A., Choi, D.-H., Brossard, D., Scheufele, D. A., Liang, X., … Dudo, A. (2012). Coverage of emerging technologies: A comparison between print and online media. *New Media & Society, 14*(6), 1039–1059. doi:10.1177/1461444812439061

Downs, A. (1957). *An economic theory of democracy*. New York, NY: Harper.

Egelhofer, J., Boyer, M., Lecheler, S., & Aaldering, L. (2022). Populist attitudes and politicians' disinformation accusations: Effects on perceptions of media and politicians. *Journal of Communication, 72*(6), 619–632. doi:10.1093/joc/jqac031

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

Fiske, S. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*(6), 889–906. https://doi.org/10.1037/0022-3514.38.6.889

Fried, I. (2023). How AI will turbocharge misinformation. *Axios.com*. Retrieved from https://www.axios.com/2023/07/10/ai-misinformation-response-measures

Gil De Zúñiga, H., & Hinsley, A. (2012). The press versus the public: What is "good journalism?" *Journalism Studies, 14*(6), 926–942. doi:10.1080/1461670X.2012.744551

Habermas, J. (2005). *Truth and justification*. Cambridge, MA: MIT Press.

Hameleers, M. (2023a). The (un)intended consequences of emphasizing the threats of mis- and disinformation. *Media and Communication, 11*(2), 5–14. doi:10.17645/mac.v11i2.6301

Hameleers, M. (2023b). Disinformation as a context-bound phenomenon: Toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. *Communication Theory, 33*(1), 1–10. doi:10.1093/ct/qtac021

Han, G., Chock, T. M., & Shoemaker, P. J. (2009). Issue familiarity and framing effects of online campaign coverage: Event perception, issue attitudes, and the 2004 Presidential Election in Taiwan. *Journalism & Mass Communication Quarterly, 86*(4), 739–755. doi:10.1177/107769900908600402

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Gulliford Press.

Hoes, E., Aitken, B., Zhang, J., Gackowski, T., & Wojcieszak, M. (2024). Prominent misinformation interventions reduce misperceptions but increase scepticism. *Nature Human Behaviour*, 1–9. Advance online publication. doi:10.1038/s41562-024-01884-x

Hogg, M. (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European Review of Social Psychology, 11*(1), 223–255. doi:10.1080/14792772043000040

Knudsen, E., Dahlberg, S., Iversen, M., Johannesson, M., & Nygaard, S. (2022). How the public understands news media trust: An open-ended approach. *Journalism, 23*(11), 2347–2363. doi:10.1177/14648849211005892

Levine, T. R. (2014). Truth-Default Theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology, 33*(4), 378–392. doi:10.1177/0261927X14535916

Lu, C., Hu, B., Li, Q., Bi, C., & Ju, X. (2023). Psychological inoculation for credibility assessment, sharing intention, and discernment of misinformation: Systematic review and meta-analysis. *Journal of Medical Internet Research, 25*(1). doi:0.2196/49255

Milmo, D., & Hern, A. (2023, May 20). Elections in UK and US at risk from AI-driven disinformation, say experts. *Guardian*. Retrieved from https://www.theguardian.com/technology/2023/may/20/elections-in-uk-and-us-at-risk-from-ai-driven-disinformation-say-experts

Newman, N., & Fletcher, R. (2017). *Bias, bullshit and lies*. Oxford, UK: Reuters Institute.

Newman, N., Fletcher, R., Eddy, K., Robertson, C. T., & Nielsen, R. K. (2023). *Reuters Institute digital news report 2023*. doi:10.60625/risj-p6es-hb13

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2022). *Reuters Institute digital news report 2022*. doi:10.60625/risj-x1gn-m549

Newman, N., Fletcher, R., Robertson, C. T., Ross Arguedas, A., & Nielsen, R. K. (2024). *Reuters Institute digital news report 2024*. doi:10.60625/risj-vy6n-4v57

Quiring, O., Ziegele, M., Schemer, C., Jackob, N., Jakobs, I., & Schultz, T. (2021). Constructive skepticism, dysfunctional cynicism? Skepticism and cynicism differently determine generalized media trust. *International Journal of Communication, 15*, 3497–3518. Retrieved from ijoc.org/index.php/ijoc/article/view/16127

Simchon, A., Edwards, M., & Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative AI. *PNAS Nexus*. doi:10.1093/pnasnexus/pgae035

Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*. doi:10.37016/mr-2020-127

Swart, J., & Broersma, M. (2023). What feels like news? Young people's perceptions of news on Instagram. *Journalism, 25*(8), 1620–1637. doi:10.1177/14648849231212737

Ternovski, J., Kalla, J., & Aronow, P. (2022). The negative consequences of informing voters about deepfakes: Evidence from two survey experiments. *Journal of Online Trust and Safety, 1*(2), 1–16. doi:10.54501/jots.v1i2.28

Tversky, A., & Kahneman, D. (1986). The framing of decisions and the evaluation of prospects. In R. Barcan Marcus, G. J. W. Dorn, & P. Weingartner (Eds.), *Studies in logic and the foundations of mathematics*. Amsterdam, the Netherlands: Elsevier. doi:10.1016/S0049-237X(09)70710-4

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society, 6*(1), 1–13. doi:10.1177/2056305120903408

Valkenburg, P., Semetko, H. A., & de Vreese, C. (1999). The effects of news frames on readers' thoughts and recall. *Communication Research, 26*(5), 550–569. doi:10.1177/009365099026005002

Van der Meer, T. G. L. A. (2018). Public frame building: The role of source usage in times of crisis. *Communication Research, 45*(6), 956–981. doi:10.1177/0093650216644027

Van der Meer, T. G. L. A., Hameleers, M., & Ohme, J. (2023). Can fighting misinformation have a negative spillover effect? How warnings for the threat of misinformation can decrease general news credibility. *Journalism Studies, 24*(6), 803–823. doi:10.1080/1461670X.2023.2187652

Zimmermann, F., & Kohring, M. (2020). Mistrust, disinforming news, and vote choice: A panel survey on the origins and consequences of believing disinformation in the 2017 German parliamentary election. *Political Communication, 37*(2), 215–237. doi:10.1080/10584609.2019.1686095