Journal of Computer-Mediated Communication, 2022, **28**:1, 1–12
https://doi.org/10.1093/jcmc/zmac024
Full-length Research Article

Fighting cheapfakes: using a digital media literacy intervention to motivate reverse search of out-of-context visual misinformation

Sijia Qian [1,*], Cuihua Shen [1] and Jingwen Zhang[1]

[1]Department of Communication, University of California, Davis, USA
*Corresponding author: Sijia Qian. Email: sjqian@ucdavis.edu

Abstract
As a significant source of misinformation, out-of-context visual misinformation refers to visuals presented in an unrelated context. This study explores whether a digital media literacy intervention that features reverse image search tools has significant effects on participants' message credibility judgment, discernment of visual misinformation, and intention of using reverse image search tools. Data were collected from a pre-registered, web-based, between-subjects experiment ($N = 905$). Results revealed a significant difference on intention of using reverse search tools among three experimental conditions: active intervention that involved both knowledge and behavior, passive intervention that involved knowledge only, and a control condition. Specifically, active intervention significantly increased intention of using reverse search tools, compared to the passive intervention and the control. Neither active nor passive intervention had an effect on credibility judgment or misinformation discernment. We discuss the implications for future digital media intervention designs and journalism practice that aim to combat visual misinformation.

Lay Summary
Social media platforms have become a fertile ground for producing and disseminating visual misinformation. One specific form of visual misinformation involves images that are presented in a false context. We designed a digital media literacy intervention that motivates and teaches users to reverse search news images when they encounter news posts on social media. We ran an online experiment comparing three groups, namely active intervention that involved providing information and instructing a reverse search practice, passive intervention that only involved providing information, and a control group where participants read irrelevant information. We found that active intervention increased individuals' motivation to use reverse image search in the future compared to passive intervention and the control group. The intervention did not have an impact on the evaluation of message credibility and misinformation discernment. Our findings suggest the potential of using a brief digital media literacy intervention to motivate future image verification practices among social media users. Social media platforms are recommended to empower users with digital literacy interventions to foster critical media consumption habits.

Keywords: visual misinformation, out-of-context visual misinformation, digital media literacy, online experiment, reverse image search, credibility, out of context

Introduction

Past research on misinformation has primarily focused on textual misinformation. As sharing and diffusion of images on social media become more commonplace, the current online media environment is no longer textual-based, but increasingly multimodal (Thomson et al., 2020). This means that misinformation can take on multiple media forms, beyond text-only presentation online. One study found that 20% of political images shared on Facebook during the months before the 2020 U.S. presidential election contained elements of misinformation (Yang et al., 2021).

Visual misinformation is defined as practices intended to manipulate (e.g., retouching), fabricate (e.g., copy, move, and splicing), and recontextualize visuals (Kasra et al., 2018). Recent studies on visual misinformation are increasingly paying attention to the production (e.g., Nguyen et al., 2019), detection (e.g., Aneja et al., 2021), and consequences of consuming visual misinformation (e.g., trust in visual media; Geddes, 2020). Visuals are studied in comparison to texts in terms of processing, effects, and debunking strategies. When visuals are presented alongside the text, they can serve as an index to reality, thus boosting the credibility of the claims (Messaris, 1997). Strategies used to address textual misinformation, such as correction (flagging and correcting misinformation after exposure) and inoculation (forewarning people against misinformation prior to exposure), have been shown to be effective in countering visual misinformation (Basol et al., 2021; Dan et al., 2021; Hameleers et al., 2020).

Among various ways to address misinformation, one understudied research area is the design and development of targeted interventions such as media literacy initiatives that can reach susceptible populations (Chou et al., 2020). Media literacy education aims to improve individuals' ability to navigate the information environment (Jeong et al., 2012). Existing research shows that media literacy interventions can be effective, although empirical evidence is sparse. For instance, studies found that digital media literacy interventions increased individuals' discernment between accurate and false news (Guess et al., 2020; Hameleers, 2022). Such an approach also has the potential to overcome the limitation of information-focused corrective strategy after the exposure, which is shown to be less effective in correcting

Associate editor: Nicole Ellison
Received: 18 March 2022. **Revised:** 29 August 2022. **Accepted:** 23 September 2022

misinformation congruent with people's prior attitudes (Nyhan & Reifler, 2010).

The current study focuses on a specific form of visual misinformation, the use of unedited and authentic images in a false context to misrepresent reality, also known as "cheapfakes." Although the creation of out-of-context visual misinformation is fairly easy and does not require technical expertise, discerning this kind of misleading content requires verification efforts from individual viewers (Fazio, 2020). Even journalists are ill-equipped to detect this type of visual misinformation, which could lead to unintentional amplification of such misinformation (Thomson et al., 2020). Equipping digital media users with the skills needed to discern out-of-context visuals is therefore crucial.

The goal of this study is to design and test the efficacy of a digital media literacy intervention on improving people's ability to discern between accurate and false multimodal online content and their intention to use image reverse search tools in the future. Few interventions have been developed to specifically address out-of-context visual misinformation, which is a more fundamental problem underlying the plague of multimodal misinformation and disinformation. We designed an intervention with practical and easy-to-use tools that individuals can routinely incorporate into their everyday media consumption. Using deficit hypothesis as the mechanism (Scherer & Pennycook, 2020) and Bandura's social cognitive theory (Bandura, 1986) as the theoretical framework in designing the digital media literacy intervention, this pre-registered study experimentally tested the effects of two intervention designs, passive (participants read an infographic about reverse image search) and active interventions (participants read the infographic and also practiced using reverse image search), in comparison to a control condition where participants read an infographic on travel information.

Theories and hypotheses

Out-of-context visual misinformation

In the academic literature, there are a few terms that are used to describe the existence of incorrect and misleading information, such as "misinformation" and "disinformation" (Wardle & Derakhshan, 2017). A key distinction between misinformation and disinformation lies in the intent. Misinformation refers to false and misleading content, while disinformation involves intentional fabrication and dissemination (Lazer et al., 2018). We choose to adopt the term misinformation since intent is not of interest in the current research and users typically cannot gauge the creator's intention of deception when encountering misinformation. Visual misinformation can pose threats to individuals, businesses, and political systems (Westerlund, 2019). On an individual level, visual misinformation can lead to misperceptions about political realities. On a population level, scholars worry that visual misinformation, especially deepfakes, can cause a loss of trust in media, authorities, and institutions (Geddes, 2020).

Out-of-context visual misinformation (also referred to as cheapfakes or visual recontextualization) remains one of the most widespread and effective forms of visual misinformation (Fazio, 2020). Specifically, it refers to presenting authentic, untouched visuals in an unrelated context or under false pretense. One study sampled COVID-19-related visual misinformation and found that 24% of the sampled misinformation contained out-of-context visuals (Brennen et al., 2021). Out-of-context visuals are also the top category of visual misinformation found in public WhatsApp groups in India, consisting of 34% of all the misinformation images (Garimella & Eckles, 2020). Note that image memes are excluded from the current discussion of out-of-context visuals, because memes are often ostensibly manipulated and modified, "typically with sarcastic or amusing undertones" (Ling et al, 2021, p. 81), and not intended to depict real events.

In the news context, out-of-context visual misinformation can be effective in influencing beliefs and actions as misinformation containing visuals is easily perceived as credible. The indexicality of visuals (i.e., true-to-life impression) adds evidence to the claims accompanied by the visuals (Messaris & Abraham, 2001). Audiences may be less suspicious of visual misinformation than textual misinformation, as images are viewed as a direct depiction of the reality. Past studies also found that visuals are more attention-grabbing than texts (Powell et al., 2015). In addition, visuals tend to be more persuasive in the sense that it could elicit a heightened emotional response (Iyer et al., 2014), and could subsequently drive attitudinal and behavioral change (e.g., von Sikorski, 2022).

Detecting visual misinformation at an early stage helps combat the engagement with and spread of visual misinformation. Identification of misinformation before it becomes viral is the first step toward deterring retransmission of misinformation (Cao et al., 2018). This is especially important for visual misinformation, as visuals tend to be more persuasive and attract greater engagement (Stenberg, 2006). One study found that tweets with images get 89% more likes and are retweeted 1.5 times more often than tweets without images (Cao et al., 2020). Although approaches to identify fake news have been developed (see Potthast et al., 2017), not all of the techniques used to detect textual misinformation can be applied to detecting visual misinformation. Specifically, detecting out-of-context misinformation poses a challenge to individuals. Luo et al. (2021) found that identifying out-of-context visuals is not an easy task, as the accuracy rate is only 66% when participants were asked to judge whether the image–caption pair is pristine or falsified. Detection is hard as the image itself is not manipulated.

Reverse image search is one of the manual detection strategies recommended by media forensics experts and organizations such as International Fact-Checking Network (Smith, 2018; Tardáguila, 2020), and is featured as the education material in this study. This approach has advantages over the correction and fact-checking method. Specifically, even though correction and fact-checking have been shown to reduce beliefs in misinformation (e.g., Walter & Murphy, 2018), they can only be done after—rather than before—the processing and distribution of the false content. Educational efforts such as brief digital media literacy interventions, on the other hand, have the potential to address the problem from the start, motivating individuals to be more critical media consumers and improving individuals' ability to preemptively spot manipulations (e.g., Basol et al., 2021). Despite these advantages, educational efforts and their evidence toward detecting visual misinformation are still lacking.

Effects of digital media literacy intervention

Media literacy, defined as the ability to critically consume media content, could be viewed as a broad umbrella term that encompasses various media-related literacy concepts depending on their respective emphases, including information, digital, and news literacy (Jones-Jang et al., 2021). For instance, news literacy could be seen as a sub-category of media literacy that focuses on news production, consumption, and contexts (Tully et al., 2020). Similarly, digital media literacy resembles the definition of media literacy, but emphasizes the digital and online nature of the media environment. Digital media literacy is defined as the skills and knowledge that individuals need to critically navigate their online media environment (Flanagin & Metzger, 2007). We focus on digital media literacy in the current study because it emphasizes both skills and knowledge and speaks more to the multimodal online environment.

Empirical evidence has documented a positive relationship between digital media literacy and the ability to discern accurate and false online content. For instance, Shen et al. (2019) found that individuals with a higher level of digital media literacy (including participants' Internet skills, photo-editing experience, and social media use) performed better in evaluating the credibility of images. Hargittai (2005) found that digital media literacy (measured as the understanding of various internet-related terms) is positively correlated with individuals' ability to locate accurate information in an online environment.

Digital media literacy interventions operate on the premises of the deficit hypothesis. The deficit hypothesis posits that people who are susceptible to misinformation are those that do not have sufficient media literacy to discern between true and false information (Scherer & Pennycook, 2020). Media literacy deficit has been long emphasized in the mass media literature. Designers of interventions teach people specific strategies to improve their media literacy and protect them from potentially harmful effects from mass media exposure. The key assumptions that underlie media literacy interventions include that (a) mass media exert direct and indirect influence on individuals and society, (b) yet many of these effects are harmful, and (c) intervention is constructed to combat these negative effects (Potter, 2010). The current digital media environment still operates on these assumptions with additional unexpected threats posed by large-scale dissemination of misinformation. In the specific context of misinformation, the harmful effects come from exposure to false information and interventions help people to acquire more critical skills for identifying misinformation. More specifically, media literacy interventions can enhance critical media skills by informing people about the information production process, enhancing knowledge about the impact that information may have on society, and teaching people concrete skills to discern between accurate information and falsehood (Hameleers, 2022).

Evidence on effects of digital media literacy interventions is mixed, partly due to large variations in study contexts and intervention designs. Some research shows that digital media literacy interventions enhance people's ability to accurately evaluate message credibility through bolstering perceived information credibility of high-quality or accurate information and decreasing the perceived credibility of low-quality or false information. A meta-analysis shows that media literacy education is generally effective in increasing knowledge and self-efficacy, and reducing risky and antisocial behaviors (Jeong et al., 2012). Guess et al. (2020) found that a digital media literacy intervention that provided tips on how to spot fake news successfully increased discernment between mainstream and false news. Specifically, in the visual misinformation context, media literacy education targeting deepfakes had a protective effect as it reduced the negative effects of disinformation messages (Hwang et al., 2021). However, the effectiveness of this approach seems to depend on the design of literacy messages. Vraga et al. (2020) did not find a significant improvement in misinformation correction when news literacy messages were presented to users. They suspected that the failure was due to the fact that the news literacy messages were not well-adapted to the online media environment. In another study, Vraga et al. (2022) found that exposure to a news literacy message increased cynicism toward information, which suggests that literacy messages have the potential to backfire.

Prior digital media literacy interventions have focused primarily on perceptual outcomes (e.g., discernment between accurate and false information). While discernment is an important outcome, it tends to be context and domain-specific, and shows short-term efficacy as a result of interventions. Most studies did not measure long-term efficacy of the intervention in fostering consistent critical media evaluations. Given this, we propose that boosting people's motivation to verify information helps to foster critical media consumption habits, which could be generalized to more diverse contexts. Therefore, intention to verify information is a valuable outcome dimension that has not been addressed extensively in prior literature. To address this gap, we include the intention to use reverse image search to verify visual information as one of the dependent variables.

In this study, we designed a digital media literacy intervention that aims to teach individuals ways to manually detect out-of-context visual misinformation. Unlike some interventions that provide tips for developing counter arguments, this intervention intends to provide practice and easy-to-use tools to identify visual misinformation. The purpose of the intervention includes (a) reminding individuals about the existence of out-of-context misinformation and the importance of vetting the source of visuals, (b) introducing a simple method, reverse image search, to examine where the visuals come from, and (c) showing individuals how to conduct reverse image search using a step-by-step instruction. The design of the intervention is consistent with past media literacy interventions (Guess et al., 2020; Hameleers, 2022; Vraga et al., 2020).

Moreover, the intervention distinguishes between active participation and passive participation. The active intervention group involves behaviors and knowledge, whereas the passive intervention group focuses solely on knowledge. Traditional interventions were typically designed to have participants passively consume information, aiming at their knowledge gains. In a recent review, Potter and Thai (2019) found that in media literacy interventions, skills and knowledge were two components that were mostly mentioned in the definition of media literacy, but half of the studies which claimed to emphasize skills did not operationalize it. More recent work has incorporated an active component where participants are prompted to proactively make decisions in a simulated environment (e.g., games). Basol et al. (2021) showed that actively playing a game was more effective in improving participants' confidence in spotting misinformation and reducing their willingness in sharing misinformation than passively consuming an infographic. Bandura's social cognitive theory (Bandura, 1986) argues that one's behavior is determined by expectancies. Based on this theory, individuals are more likely to adopt a behavior when they believe they are personally capable of doing so (i.e., efficacy expectations). When individuals actively participate by practicing reverse image search as part of the intervention, they are more likely to get motivated and gain skills in the process, which could lead to the adoption of using reverse image search to verify the source of images when they encounter news online. Based on the above discussion, the following hypothesis is proposed:

H1: Active intervention will increase perceived credibility of accurately attributed visual posts (H1a), reduce perceived credibility of misattributed visual posts (H1b), improve the ability to distinguish between accurately attributed and misattributed visual posts (H1c), and increase the intention to use the reverse image search tools in the future compared to passive intervention and control (H1d).

Past research shows that media literacy is positively associated with skepticism. Specifically, individuals who are more news literate are more skeptical toward online information in general (Vraga & Tully, 2021). When it comes to visuals, the level of digital media literacy is also positively associated with skepticism of visual information overall, regardless of its veracity (Shen et al., 2019). In the current study context, a digital media literacy intervention might have a stronger effect on discernment for individuals who are already more digital media literate, as the intervention will help decrease their skepticism. While some scholars argue that a visual-specific literacy is needed in identifying visual manipulations (Lazard et al., 2020), general media literacy is found to be more effective in reducing the persuasiveness of and intention to share visual disinformation than deepfake-specific literacy (Hwang et al., 2021). Similarly, Barari et al. (2021) showed that having a broad understanding of the digital environment increases deepfake detection. On the other hand, it is also likely that the intervention will have a stronger effect among participants with lower baseline digital media literacy as their ability to correctly evaluate message credibility is much lower than those with higher baseline digital media literacy. We deem discernment as the primary outcome as it is the most common measurement used to assess the efficacy of media literacy interventions in helping people discern facts from falsehood. The following research question is proposed to examine the moderation effect of digital media literacy on the relationship between the intervention and discernment:

RQ1: Will digital media literacy moderate the effect of intervention on the ability to distinguish between accurately attributed and misattributed visual posts?

Previous literature has established that individuals' political leaning also plays a role in the evaluation of message credibility and belief in misinformation, even after being exposed to a correction of the misinformation (Hameleers et al., 2020). Social media use is also found to be associated with critical consumption of media content (Xiao et al., 2021). Moreover, studies show that a low level of trust in media will make individuals more skeptical toward news and mediated information (e.g., Tsfati, 2003). Given that the study examines evaluations of visual news posts and promotes the use of image reverse searching tools, visual literacy is likely to influence these outcomes (Brumberger, 2011). Therefore, we include political leaning, social media use, trust in media, and visual literacy as covariates.

Method

Study design

The study is preregistered in OSF[1] and is designed as a between-subjects experiment embedded in an online Qualtrics survey. We used a single-factor design with random assignment to one of three groups (active intervention, passive intervention, and control). Active intervention included an educational infographic and a task to practice reverse image search. Passive intervention only included an education infographic that is the same as in the active intervention. The control condition involved a travel-related infographic designed in a similar format.

Digital media literacy intervention

Consistent with the conceptualization of digital media literacy, the design of educational messages focuses on both literacy and skill (see Figure 1 for the full design details). In terms of the literacy component, the educational infographic gave an introduction to reverse image search (e.g., "Reverse image search is when you use an image—instead of a keyword—to search the web") and discussed the importance of vetting the source of images in online news (e.g., "Verifying the source of the news image could help stop spreading misleading news posts."). As for the skill component, the infographic provided step-by-step instructions on using reverse image search. The infographic was designed based on manuals from Common Sense Education (https://www.commonsense.org/education/tips-resources). Since reverse image search is conducted slightly differently based on the browser one is using, tips tailored to each browser were also provided.

In the active intervention group, participants were also instructed to practice reverse image search by finding a caption that best matches the image provided (i.e., an image featuring the Deep Space Station 43 in Canberra, Australia). The captions submitted by participants were checked to see if they matched the actual reverse image search results. Captions that were too general (e.g., "satellite") or irrelevant (e.g., "none just a pic") were considered as failing the task. This practice aimed to provide an opportunity for the participants to actively participate in reverse image search and practice the steps they learned from the intervention infographic. Participants did not receive any feedback for this task during the experiment, and those who failed the task were excluded from analysis.

Stimuli

A total of eight visual posts were created based on four images (4 images × 2 captions: accurately captioned vs. misattributed). The four unedited images depicted (a) a park covered in litter, (b) the Black Lives Matter (BLM) bus, (c) women wearing masks, and (d) vegan food on grocery store shelves, respectively. Each visual post contained one image and accompanying text caption that was sampled from fact-checking articles on Snopes.com. The text captions either accurately described the events captured in the image or describe a completely unrelated scenario (thus being out-of-context). The captions covered diverse sociopolitical issues such as BLM, climate change, and vegan food. All posts were

Figure 1. Digital media literacy infographic used in active and passive intervention groups.

designed using a generic social media post format without platform logos (see Figure 2 for examples), considering that individuals likely consume news using different social media sites. To avoid biased processing, the source of the post (i.e., sender) was presented using gender-neutral names (i.e., Alex Smith) and the profile picture used object-centered images (i.e., flowers). At the bottom of each post, numbers of likes and comments were shown and were held constant across posts. Regardless of experimental assignment, each participant saw all four images, two of which were accurately captioned and two were misattributed.

Sample and procedure

Participants were recruited from the online subject recruitment platform Prolific ($N = 905$, 49.6% female with the mean age of 44.5, $SD = 16.1$). Each participant was paid \$1.10 for their time. To be eligible for the study, individuals had to be at least 18 years old, read English, and reside in the U.S. There were no significant differences in demographics between the experimental groups, including age, sex, education levels, and household income.

Participants were required to take the survey on a desktop or laptop device, but not a mobile device, because reverse image search is much more difficult to carry out when using a mobile device. Figure 3 presents a flow diagram of the study procedure. Participants first answered questions regarding digital media literacy, visual literacy, and trust in media. Then, they were randomly assigned to one of three groups: an active intervention, a passive intervention, or a control group. In the active intervention group, participants were shown a digital media literacy education infographic (see Figure 1) and were instructed to practice reverse image search by finding a caption that best matches the image provided. In the passive intervention group, they were shown an educational infographic that was the same as in the active intervention group, but did not receive additional instructions on using the reverse image search tool. Participants in the control group read a travel-related infographic designed in a similar format. In all groups, participants were instructed to read the infographic for at least 30 s before proceeding to the next page. Participants in the active and passive intervention groups also answered an attention check question after receiving the intervention treatment.

After, all participants read an introductory text that explained they were going to read four visual news posts taken directly from various social media sites. All participants saw all four images, two of which were misattributed and two were accurately attributed. There were six possible combinations of the eight visual posts (4 images $\times$ 2 captions: accurately captioned vs. misattributed) participants were randomly assigned to view. The visual news posts were presented to the participants in a random order. Participants were required to view each post for at least 10 s before proceeding to the next. After each visual news post, they immediately rated the perceived credibility of the post, before viewing the next one. Lastly, they answered questions about intention of using reverse image search in the future, political ideology, and other demographic questions.

Measures

Dependent variables

Perceived credibility

Participants rated the credibility of each visual post using a five-item scale ($1 =$ *Strongly Disagree*; $5 =$ *Strongly Agree*).

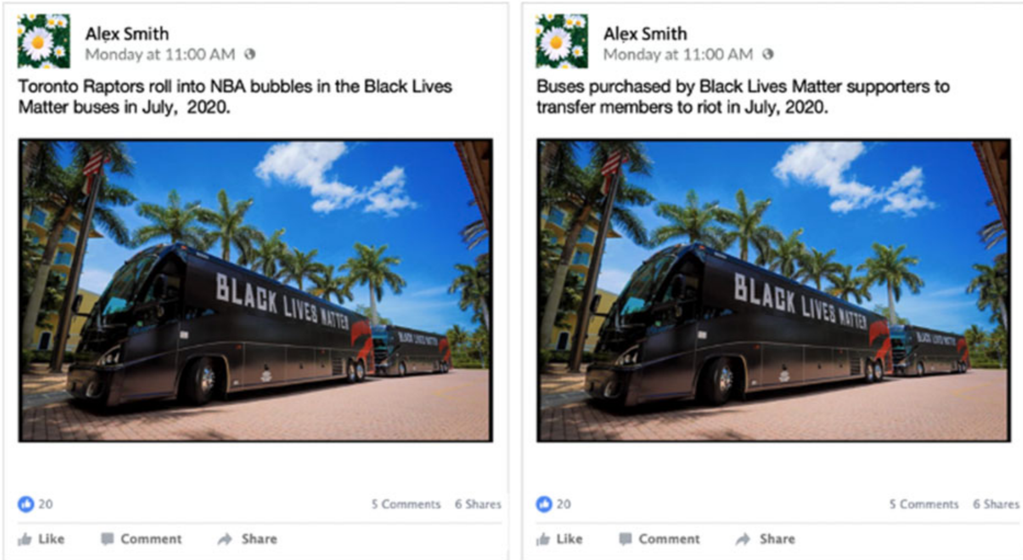

Figure 2. Left: Example misattributed visual post (text caption is unrelated to the original image context). Right: Example accurately attributed visual post (text caption is consistent with the original image context).

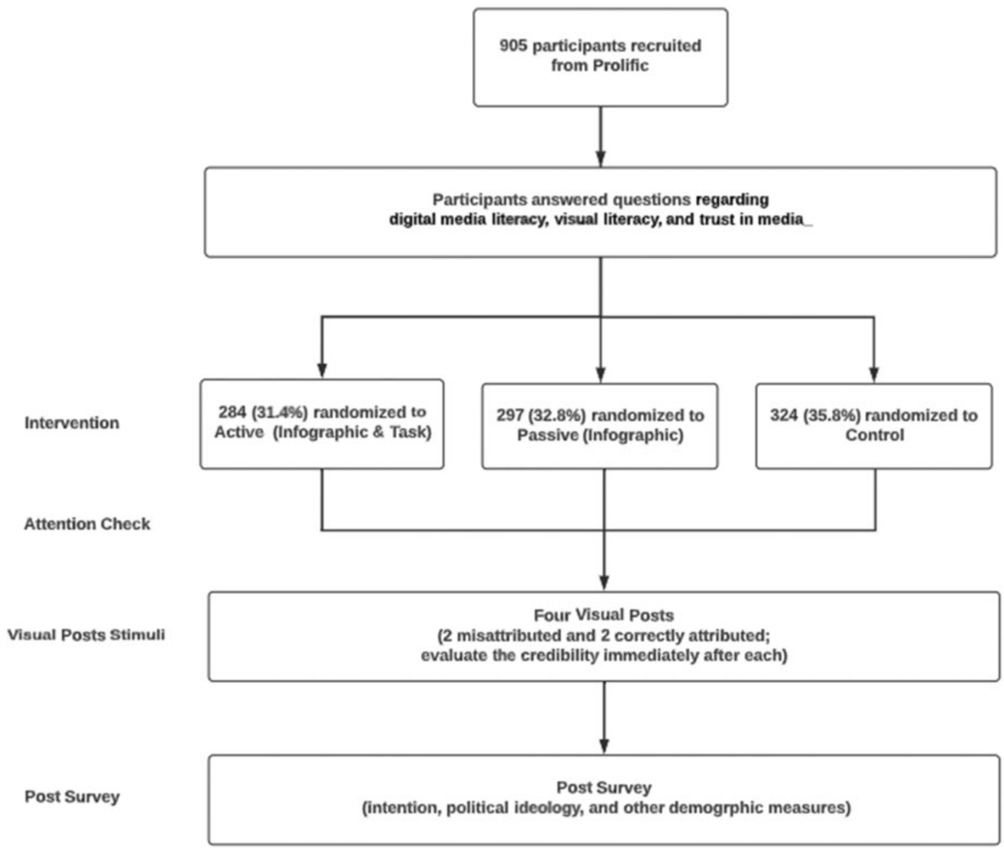

Figure 3. Flow diagram of study procedure.

The items include (a) “The post was credible,” (b) “The post was accurate,” (c) “The post reflected reality,” (d) “The post contained no falsehoods,” and (e) “The post was fake” (reverse coded). The five items were averaged into a composite measure and showed high reliability (Cronbach’s alpha = 0.93). The measurement was adapted from existing measures of message credibility with a focus on perceived accuracy (e.g., Hameleers, 2022). Perceived credibility of accurately attributed visual posts was calculated as the average of the credibility of the two accurately attributed visual posts participants viewed. Perceived credibility of misattributed visual posts was calculated in a similar way.

Discernment

Discernment was calculated as the mean difference in perceived credibility between all posts with accurate captions and posts with misattributed captions viewed at the respondent level, consistent with Guess et al. (2020).

Intention to use the reverse image search in the future

Participants were asked how likely they would use reverse image search tools (e.g., Google image search) when browsing through visual posts on social media in the future on a 5-point scale (1 = *Extremely Unlikely*; 5 = *Extremely Likely*).

Moderator

Digital media literacy

Participants indicated their perceived familiarity with nine Internet-related items (1 = *No Understanding; 5 = Full Understanding*). The items include advanced search, tagging, PDF, spyware, Wiki, JPG, cache, malware, and phishing. The scale is adopted from Hargittai (2005). The nine items showed excellent reliability (Cronbach's alpha = 0.92) and were averaged into a composite measure.

Covariates

Political ideology

Political ideology was measured using two items. Participants indicated their ideology in social issues and economic issues ranging from very liberal to very conservative in a 7-point scale (Akin et al., 2019). The two items were averaged into a measure of political ideology ($r = 0.85$). Political ideology was significantly correlated with party membership (i.e., Democratic participants were more liberal and Republican participants were more conservative; $r = 0.79$, $p < .001$).

Visual literacy

Participants rated their levels of skill with visual creation and digital creation tools, including photography, image manipulation, digital illustration, and reverse image search (Brumberger, 2011; Lazard et al., 2020) using a 5-point scale (1 = *No Experience*; 5 = *Expert/Professional*). These items showed satisfactory reliability and were combined into one measure (Cronbach's alpha = 0.75).

Trust in media was measured by asking participants how much they agree with the statements that they trust mass media and social media respectively in reporting news fully, accurately, and fairly (1 = *Strongly Disagree*; 5 = *Strongly Agree*; Guess et al., 2020). The two items were significantly correlated ($r = .35$, $p < .001$). Since the two items tap into different aspects of media trust, they were treated as individual items in the analysis, i.e., trust in mass media and trust in social media.

Social media use

Participants indicated the frequency of using seven social media platforms, including Facebook, Twitter, Instagram, Snapchat, YouTube, Reddit, and TikTok (1 = *Never*; 5 = *Always*). They were also given the choice to supplement one other social media site and indicate their level of use. We chose the cutoff point as the middle value (i.e., 3) and counted the number of social media sites associated with a value of 3 and above.

Age, sex, education, and income served as the primary control variables in the analyses. Age was measured as a continuous variable. Sex was a binary variable with female as the reference group. Education was measured as an ordinal variable by asking respondents the highest degree or level of education they have completed (*Median* = 6, *SD* = 1.74). The median value corresponded to "Bachelor's Degree." Household income was measured as an ordinal variable. Categories ranged from "less than US$10,000" (coded as 1) to "US$100,000 or more" (coded as 11; $M = 6.53$, $SD = 3.32$). The median value for income was 6, indicating "US$50,000 to US$59,999." Correlations of all the continuous variables involved in this study were reported in Table 1, with their means and standard deviations.

Attention check

Participants in the active and passive intervention groups were asked to select the statement that matched the definition of reverse image search. The options included "A reverse image search is when you use an image—instead of a keyword—to search the web" (correct answer), "A reverse image search is using keywords to search for an image," "A reverse image search is complicated and not worth trying," and "Not sure/Don't know." Participants who chose the other options were removed from the analysis.

Analytical strategy

We used both Analysis of Variance (ANOVA) and Ordinary Least Square (OLS) regression models to test H1a to H1d. Four one-way ANOVA models were run to test the main effects of

Table 1. Correlation of continuous variables ($N = 905$)

	Mean	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1 Age	44.53	16.05												
2 Education	5.27	1.74	.12***											
3 Income	6.53	3.32	−.01	.34***										
4 Media use	3.98	1.75	−.34***	.03	.07**									
5 Trust in mass media	2.87	1.26	−.01	.10**	.04	.14***								
6 Trust in social media	2.15	1.10	−.17***	−.004	.02	.30***	.35***							
7 Digital media literacy	3.95	0.84	−.08	.03	.06*	.20***	−.04	−.05						
8 Visual literacy	2.21	0.78	−.18***	.09**	.12**	.31***	.07	.16***	.44***					
9 Political leaning	3.17	1.82	.19***	−.05	.08*	−.09**	−.39***	−.07*	−.02	−.03				
10 Perceived credibility of accurately attributed visual posts	2.99	0.75	−.06	−.01	.06	.12***	.02	.19***	−.01	.06	.09**			
11 Perceived credibility of misattributed visual posts	2.95	0.82	.07*	−.12**	−.07	−.02	−.10**	.13***	−.07	−.03	.28***	.29***		
12 Discernment	0.04	0.92	−.11***	.10*	.10**	.12**	.11**	.03	.05	.07*	−.17***	.54***	−.65***	
13 Intention to use reserve image search	3.39	1.30	−.10**	−.05	.004	.16***	.13***	.12***	.12***	.25***	−.13***	−.02	−.11**	.08*

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

intervention with perceived credibility of accurately attributed visual posts, perceived credibility of misattributed visual posts, discernment, and behavioral intention respectively as dependent variables. Pairwise t tests with Bonferroni corrections were also conducted to compare across three experimental groups. We also ran four OLS regression models with intervention as two categorical variables (active vs. control and passive vs. control) and the same set of dependent variables to further test the effects of intervention above and beyond covariates. These covariates include demographic variables (i.e., age, sex, education, and income), social media use, political ideology, digital media literacy, visual literacy, and trust in mass media and social media. The regression models were also run using active intervention as the reference group to compare the effects between active and passive intervention groups. Lastly, the moderation effect of digital media literacy was tested by an extra sum of squares F test comparing the model with the interaction term to the model above without the interaction term (RQ1).

Results

Out of 608 participants in the intervention groups, 597 participants passed the attention-check. Those participants that did not pass the attention check question were removed from the analysis (three in the active intervention group and eight in the passive intervention group). For participants in the active intervention group, 14 participants failed to complete the relevant caption task (i.e., their captions were either too general or totally irrelevant). They were also removed from the analysis. As a result, the total number of participants included in the analysis below is 880, including 267 participants in the active intervention group, 289 participants in the passive intervention group, and 324 participants in the control group. They did not differ significantly in terms of age, sex, education levels, and household income.

To test the main effects of intervention proposed in H1a to H1d, four one-way ANOVA tests were conducted. H1a predicted that active intervention would increase the perceived credibility of accurately attributed visual posts compared to passive intervention and the control. H1a was not supported (F (2, 877) = 0.73, p = .481). Post-hoc pairwise comparison did not reveal any significant differences across each comparison group. H1b tested the effects of intervention on the perceived credibility of misattributed visual posts. We did not find a significant main effect supporting H1b (F (2, 877) = 2.10, p = .123). H1c predicted the active intervention would increase discernment between accurately attributed and misattributed visual posts. H1c was not supported (F (2, 877) = 1.37, p = .254). Lastly, H1d hypothesized that active intervention would increase participants' intention to use the reverse image search in the future. We found a significant main effect of intervention (F (2, 877) = 32.52, p < .001). Looking at specific comparisons, active intervention (M = 3.78, SE = 1.16, p <.001) and passive intervention (M = 3.51, SE = 1.26, p <.001) significantly increased intention compared to the control group (M = 2.97, SE = 1.34). Participants in the active intervention group displayed higher levels of intention than those in the passive intervention group (p = .026).

To further examine the predictors of individuals' credibility judgment and behavioral intentions, we ran four OLS regression models including a group of covariates that were shown to be significantly correlated with one or more of the outcomes (see results in Table 2).[2] While controlling for other covariates, intervention did not significantly impact participants' perceived credibility of accurately attributed and misattributed visual posts and discernment between accurately attributed and misattributed visual posts. The existence of an intervention did increase individual's intention to practice reverse image search in the future compared to the control group (active intervention: B = 0.75, t = 7.30, p <.001; passive intervention: B = 0.53, t = 5.33, p <.001). Moreover, active intervention significantly increased intention compared to passive intervention (B = 0.22, t = 2.04, p = .042).

RQ1 asked whether digital media literacy would moderate the effects of intervention on discernment. We conducted an extra sum of squares F test comparing the model with an interaction term between digital media literacy and intervention and the model without the interaction term (not shown in Table 2). The model comparisons revealed that digital media literacy did not moderate the relationship between intervention and discernment (F (2,831) = 0.994, p = .371).

Table 2. OLS regression models (using control group as the reference group; N = 844)

	Perceived credibility of accurately attributed visual posts B (SE)	Perceived credibility of misattributed visual posts B (SE)	Discernment B (SE)	Intention to use reverse image search B (SE)
Intercept	2.615*** (.191)	2.727*** (.203)	−0.112 (.238)	2.102*** (.316)
Age	−0.001 (.002)	0.004 (.002)	−0.004* (.002)	−0.0001 (.003)
Sex (Male = 1)	0.091 (.051)	0.057 (.054)	0.033 (.064)	−0.038 (.085)
Education	−0.016 (.016)	−0.039* (.017)	0.024 (.020)	−0.047 (.026)
Income	0.009 (.008)	−0.014 (.009)	0.023* (.010)	0.005 (.013)
Social media use	0.035* (.017)	−0.001 (.017)	0.036 (.021)	0.062* (.027)
Political ideology (Conservative = High)	0.042** (.015)	0.114*** (.016)	−0.072*** (.019)	−0.053* (.026)
Digital media literacy	−0.018 (.034)	−0.037 (.036)	0.018 (.043)	0.014 (.057)
Visual literacy	0.004 (.038)	−0.007 (.040)	0.011 (.047)	0.339*** (.063)
Trust in mass media	−0.010 (.024)	−0.029 (.025)	0.019 (.029)	0.066 (.039)
Trust in social media	0.103*** (.026)	0.132*** (.028)	−0.030 (.032)	0.025 (.043)
Intervention_active	0.018 (.062)	−0.090 (.066)	0.108 (.077)	0.748*** (.103)
Intervention_passive	−0.070 (.060)	−0.113 (.064)	0.043 (.075)	0.533*** (.100)
Adjusted R^2	0.055	0.126	0.059	0.140
F	(12, 831) = 4.05***	(12, 831) = 9.98***	(12, 831) = 4.35***	(12, 831) = 12.15***

Note. Thirty-six observations deleted due to missing values in age, sex, and income.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

This study design did not capture whether participants actually performed reverse image search when evaluating visual posts. We compared the time spent on evaluating visual posts among the three conditions as a proxy measure. One-way ANOVA test shows that there was a significant difference among groups (F (2, 891) = 12.08, p <.001). Specifically, participants spent significantly longer time in the active intervention group when evaluating the four visual posts ($M = 29.87$) than in the passive ($M = 22.80$) and the control group ($M = 20.16$), while the passive intervention group did not differ significantly from the control group.

Discussion

Out-of-context visual misinformation, also known as cheapfakes or visual recontextualization, refers to the practice of using authentic and untouched images in an unrelated context to misrepresent reality. Through a preregistered survey experiment, the current study examined the effects of a digital media literacy intervention on participants' visual post-credibility evaluation, discernment, and related behavioral intentions. Findings suggested that, while exposure to the intervention did not influence the ability to identify accurately attributed and misattributed visual posts, it significantly increased participants' intention of using reverse image search in the future, which is one of the best visual misinformation detection methods at the moment (Fazio, 2020; Smith, 2018). We did not find evidence that participants' existing digital media literacy moderated the relationship between intervention and credibility discernment.

Regarding the effects of intervention, results showed that intervention was effective in increasing the intention of using reverse image search in the future. One important aspect of media literacy is participants' ability to adapt to new technologies (Coiro et al., 2014). The intervention was successful in equipping people with the skills they need and motivating people to practice visual misinformation search. This is crucial as the baseline use of reverse image search is possibly low. Although there is no population-level study on this to our knowledge, in our study sample, more than one-quarter of the participants (28.5%) reported having no experience of using reverse image search tools, and only half of the participants reported being slightly or somewhat skilled with reverse image search tools (49.2%). Moreover, the theory of reasoned action argues that the primary predictor of behavior is one's intention to perform that behavior (Ajzen & Fishbein, 1980). Increasing people's intention to use reverse image search means that there is a high likelihood that people will perform reverse image search in the future. However, whether intention could translate to behavioral use will need more empirical tests by future longitudinal studies. While we did not measure whether reverse image search was actually performed by participants to judge the credibility of the visual posts, we showed that participants in the active and passive intervention groups spent significantly more time evaluating the visual posts compared to the control. This suggests that participants were indeed paying more attention to visual posts after they read the infographic and practiced the skills themselves. This finding highlights the importance of involving both knowledge and behavior in the intervention design.

Intervention did not influence perceived credibility or discernment. One possible explanation is that the eight visual posts based on four images (both misattributed and accurately attributed) did not have similar levels of baseline credibility, possibly due to participants' prior exposure to related news. To assess baseline credibility, we conducted independent sample t-tests within the control group. Out of the four images, two images had similar levels of perceived credibility between accurately attributed and misattributed posts (i.e., a park covered in litter and vegan food on grocery store shelves; see Table B1 in the Supplementary Material). In contrast, participants viewed the accurately attributed BLM bus post as significantly more credible than its misattributed counterpart, and the misattributed post featuring mask-wearing women significantly more credible than its accurately attributed counterpart. Even though each participant saw all four images, they were randomly assigned to view two accurately attributed and two misattributed posts, and we averaged their credibility ratings across image posts to construct the measures. Therefore, the effects of intervention might have been washed out because of the different combinations of image contexts. We then performed the same analysis within each image context and caption individually (see Appendix C in the Supplementary Material). Results showed that across all eight visual posts, both active and passive intentions significantly increased the intent to use reverse image search compared to the control, which is consistent with the result of our main analysis using the entire sample. Yet, the relationships between interventions and credibility perceptions were mostly insignificant, likely due to insufficient power. And in the few cases where they were significant, their directions were somewhat inconsistent (see Tables C1(a), C2(a), and C4(b) in the Supplementary Material). These results suggest that the intervention effects on credibility judgment may not be consistent across image contexts and future research needs to test more categories of contexts to examine in detail the boundaries of intervention effects based on people's prior associations with issue contexts.

It should be noted that, although there was no significant improvement in perceived credibility of accurately attributed visual posts, the intervention did not lower the credibility of accurately attributed visual posts. One study found an unintended spillover effect of using general warnings. Presenting general warning messages (e.g., reminding people of the existence of misinformation) could diminish the accuracy perception for accurate news and result in an increased level of distrust in the media (Clayton et al., 2020). In this regard, the current intervention did not yield such unintended effect.

Digital media literacy did not moderate the relationship between intervention and discernment. This suggests that the effects of intervention might not depend on prior literacy. Participants in the study also reported a somewhat high level of digital media literacy ($M = 3.97$ on a 5-point scale), which may also contribute to the null effect.

As a covariate, digital media literacy did not relate to identifying fake news or discerning between accurate and fake news. This is in line with recent empirical evidence from Jones-Jang et al. (2021), which found that among different types of literacies, only information literacy increased the likelihood of identifying fake news. Consistent with previous literature, trust in social media news boosted the perceived credibility of accurately attributed visual posts as well as misattributed visual posts. Political ideology played a key influence in perceived message credibility, discernment, and behavioral intentions. Future study should test how digital media literacy interventions could be adapted to audiences with varying levels of trust in social media and different political ideology.

Several limitations should be considered. First, the study only examined the immediate effects of a digital media intervention. A multi-wave panel design might be needed to test for long-term and continued influence. The results showed that intervention influenced behavioral intention of using reverse image search, which might subsequently lead to increased discernment. We recommend future research to test the long-term effects of digital media literacy interventions and examine whether intention could translate to behaviors and subsequently influence discernment.

Second, the influence of image selection on results was stronger than we expected. Message credibility had a big variance among four news contexts in our study and between the accurately attributed and misattributed visual posts in the baseline control condition (see Table B1 in the Supplemental Material), and also in the other two experimental conditions (see Tables B2 and B3 in the Supplemental Material). Future research should consider doing more rigorous pretesting to ensure that mis- and accurately-attributed visual posts have equivalent levels of baseline credibility within and across image contexts. This could help eliminate image-specific effects on the results.

Lastly, the current reverse image search interface is not mobile-friendly. Participants typically need to save the image to their local device and upload it to the image search site or to download a specific app. Therefore, to avoid these hurdles, this study limited participation on desktop devices only, potentially limiting the study's validity and real-world implications as news consumption increasingly happens on mobile devices. One potential solution to this problem is photo provenance information in news posts. For instance, one project aims to provide original details of the photo and show additional photos of the scene to help individuals critically evaluate the information (Koren, 2019). Future research should examine the impacts of automatic reverse image search and how they may contribute to correct discernment of visual posts on social media.

Conclusion

This study examines the efficacy of a digital media literacy intervention on combating out-of-context visual misinformation. While there were no significant effects of the intervention on the evaluation of message credibility or misinformation discernment, the intervention was successful in increasing participants' intention of using reverse research in the future. This shows that our simple intervention could potentially enable individual social media users to integrate image verification practices in their daily online news consumption. The comparison between active and passive interventions suggests media literacy interventions should not only focus on providing information, but also consider incorporating a more concrete skill/behavior component through actual practices. The findings also suggest that social media platforms should consider including more metadata and contextual information when presenting visuals to assist visual information verification. Platforms, fact-checking agencies, and educators could join forces to design media literacy interventions on technology platforms to encourage specific user groups to verify information before consumption or dissemination behaviors occur.

Data availability

The data collected by this study will be shared on reasonable request to the corresponding author.

Supplementary material

Supplementary material is available at *Journal of Computer-mediated Communication* online.

Notes

1. https://osf.io/ra3zd/?view_only=518b6d3334f640a68c8e1a01da29d221
2. We ran the same analyses on the full sample and the results remained largely the same (see Table A1 in the Supplemental Material).

References

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Prentice Hall. [Database]

Akin, H., Yeo, S. K., Wirz, C. D., Scheufele, D. A., Brossard, D., Xenos, M. A., & Corley, E. A. (2019). Are attitudes toward labeling nano products linked to attitudes toward GMO? Exploring a potential 'spillover' effect for attitudes toward controversial technologies. *Journal of Responsible Innovation*, *6*(1), 50–74. https://doi.org/10.1080/23299460.2018.1495026

Aneja, S., Bregler, C., & Nießner, M. (2021). Catching out-of-context misinformation with self-supervised learning. *arXiv preprint*. https://doi.org/10.48550/arXiv.2101.06278

Bandura, A. (1986). *Social foundations of thought and action*. Prentice Hall.

Barari, S., Lucas, C., & Munger, K. (2021). Political deepfakes are as credible as other fake media and (sometimes) real media. *OSF preprints*. https://doi.org/10.31219/osf.io/cdfh3

Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W., & van der Linden, S. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*, *8*(1), 1–21. https://doi.org/10.1177/20539517211013868

Brennen, J. S., Simon, F. M., & Nielsen, R. K. (2021). Beyond (mis)representation: Visuals in COVID-19 misinformation. *The International Journal of Press/Politics*, *26*(1), 277–299. https://doi.org/10.1177/1940161220964780

Brumberger, E. (2011). Visual literacy and the digital native: An examination of the millennial learner. *Journal of Visual Literacy*, *30*(1), 19–47. https://doi.org/10.1080/23796529.2011.11674683

Cao, J., Guo, J., Li, X., Jin, Z., Guo, H., & Li, J. (2018). Automatic rumor detection on microblogs: A survey. *arXiv preprint*. https://doi.org/10.48550/arXiv.1807.03505

Cao, J., Qi, P., Sheng, Q., Yang, T., Guo, J., & Li, J. (2020). Exploring the role of visual content in fake news detection. *Disinformation, Misinformation, and Fake News in Social Media*, 141–161. https://doi.org/10.1007/978-3-030-42699-6

Cao, J., Qi, P., Sheng, Q., Yang, T., Guo, J., & Li, J. (2020). Exploring the role of visual content in fake news detection. In K. Shu, S. Wang, D. Lee, & H. Liu (Eds.), *Disinformation, misinformation, and fake news in social media. Lecture Notes in Social Networks*. Springer. https://doi.org/10.1007/978-3-030-42699-6_8

Chou, W.-Y. S., Gaysynsky, A., & Cappella, J. N. (2020). Where we go from here: Health misinformation on social media. *American Journal of Public Health*, *110*(S3), S273–S275. https://doi.org/10.2105/AJPH.2020.305905

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior, 42*(4), 1073–1095. https://doi.org/10.1007/s11109-019-09533-0

Coiro, J., Knobel, M., Lankshear, C., & Leu, D. J. (2014). *Handbook of research on new literacies*. Routledge.

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

Fazio, L. (2020, February 14). *Out-of-context photos are a powerful low-tech form of misinformation*. The Conversation. http://theconversation.com/out-of-context-photos-are-a-powerful-low-tech-form-of-misinformation-129959

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, *9*(2), 319–342. https://doi.org/10.1177/1461444807075015

Garimella, K., & Eckles, D. (2020). Images and misinformation in political groups: Evidence from WhatsApp in India. *arXiv preprint*. https://doi.org/10.48550/arXiv.2005.09784

Geddes, K. (2020). Ocularcentrism and deepfakes: Should seeing be believing? *Fordham Intellectual Property, Media & Entertainment Law Journal*, *31*(4), 1042.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

Hameleers, M. (2022). Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society*, *25*(1), 110–126. https://doi.org/10.1080/1369118X.2020.1764603

Hameleers, M., Powell, T. E., Van Der Meer, T. G., & Bos, L. (2020). A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. *Political Communication*, *37*(2), 281–301. https://doi.org/10.1080/10584609.2019.1674979

Hargittai, E. (2005). Survey measures of web-oriented digital literacy. *Social Science Computer Review*, *23*(3), 371–379. https://doi.org/10.1177/0894439305275911

Hwang, Y., Ryu, J. Y., & Jeong, S.-H. (2021). Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 188–193. https://doi.org/10.1089/cyber.2020.0174

Iyer, A., Webster, J., Hornsey, M. J., & Vanman, E. J. (2014). Understanding the power of the picture: The effect of image content on emotional and political responses to terrorism. *Journal of Applied Social Psychology*, *44*(7), 511–521. https://doi.org/10.1111/jasp.12243

Jeong, S.-H., Cho, H., & Hwang, Y. (2012). Media literacy interventions: A meta-analytic review. *Journal of Communication*, *62*(3), 454–472. https://doi.org/10.1111/j.1460-2466.2012.01643.x

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, *65*(2), 371–388. https://doi.org/10.1177/0002764219869406

Kasra, M., Shen, C., & O'Brien, J. (2018). Seeing is believing: How people fail to identify fake images on the web. In *Proceedings of the 27th ACM Conference on Human Factors in Computing Systems* (CHI 2018), Montreal, Canada. https://doi.org/10.1145/3170427.3188604

Koren, S. (2019, July 23). *Introducing the News Provenance Project*. NYT Open. https://open.nytimes.com/introducing-the-news-provenance-project-723dbaf07c44

Lazard, A. J., Bock, M. A., & Mackert, M. S. (2020). Impact of photo manipulation and visual literacy on consumers' responses to persuasive communication. *Journal of Visual Literacy*, *39*(2), 90–110. https://doi.org/10.1080/1051144X.2020.1737907

Lazer, D., Baum, M., Benkler, J., Berinsky, A., Greenhill, K., Metzger, M., . . . Zittrain, J. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aar4533

Ling, C., AbuHilal, I., Blackburn, J., De Cristofaro, E., Zannettou, S., & Stringhini, G. (2021). Dissecting the meme magic: Understanding indicators of virality in image memes. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–24. https://doi.org/10.1145/3449155

Luo, G., Darrell, T., & Rohrbach, A. (2021). Newsclippings: Automatic generation of out-of-context multimodal media. *arXiv preprint*. https://doi.org/10.48550/arXiv.2104.05893

Messaris, P. (1997). *Visual persuasion: The role of images in advertising*. Sage. https://doi.org/10.4135/9781452233444

Messaris, P., & Abraham, L. (2001). The role of images in framing news stories. In S. D. Reese, O. H. Gandy Jr., & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 215–226). Lawrence Erlbaum Associates Publishers.

Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., & Nahavandi, S. (2019). Deep learning for deepfakes creation and detection: A survey. *arXiv preprint*. https://doi.org/10.48550/arXiv.1909.11573

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Potter, J. (2010). The state of media literacy. *Journal of Broadcasting & Electronic Media*, *54*(4), 675–696. https://doi.org/10.1080/08838151.2011.521462

Potter, J., & Thai, C. (2019). Reviewing media literacy intervention studies for validity. *Review of Communication Research*, *7*, 38–66. https://doi.org/10.12840/ISSN.2255-4165.018

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint*. https://doi.org/10.48550/arXiv.1702.05638

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2015). A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication*, *65*(6), 997–1017. https://doi.org/10.1111/jcom.12184

Scherer, L. D., & Pennycook, G. (2020). Who is susceptible to online health misinformation? *American Journal of Public Health*, *110*(S3), S276–S277. https://doi.org/10.2105/AJPH.2020.305908

Shen, C., Kasra, M., Pan, W., Bassett, G. A., Malloch, Y., & O'Brien, J. F. (2019). Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. *New Media & Society*, *21*(2), 438–463. https://doi.org/10.1177/1461444818799526

Smith, B. (2018, February 10). Fake News, hoax Images: How to spot a digitally altered photo from the real deal. *ABC News*. https://www.abc.net.au/news/science/2018-02-11/fake-news-hoax-images-digitallyaltered-photos-photoshop/9405776

Stenberg, G. (2006). Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology*, *18*(6), 813–847. https://doi.org/10.1080/09541440500412361

Tardáguila, C. (2020, January 7). How to use your phone to spot fake images surrounding the U.S.-Iran conflict. *Poynter*. https://www.poynter.org/fact-checking/2020/how-to-use-your-phone-to-spot-fake-images-surrounding-the-u-s-iran-conflict/

Thomson, T. J., Angus, D., Dootson, P., Hurcombe, E., & Smith, A. (2020). Visual mis/disinformation in journalism and public communications: Current verification practices, challenges, and future opportunities. *Journalism Practice*, *16*(5), 938–962. https://doi.org/10.1080/17512786.2020.1832139

Tsfati, Y. (2003). Does audience skepticism of the media matter in agenda setting? *Journal of Broadcasting & Electronic Media*, *47*(2), 157–176. https://doi.org/10.1207/s15506878jobem4702_1

Tully, M., Vraga, E. K., & Bode, L. (2020). Designing and testing news literacy messages for social media. *Mass Communication and Society*, *23*(1), 22–46. https://doi.org/10.1080/15205436.2019.1604970

von Sikorski, C. (2022). Visual polarisation: Examining the interplay of visual cues and media trust on the evaluation of political candidates. *Journalism*, *23*(9), 1900–1918. https://doi.org/10.1177/1464884920987680

Vraga, E. K., Bode, L., & Tully, M. (2020). Creating news literacy messages to enhance expert corrections of misinformation on Twitter.

Communication Research, *49*(2), 245–267. https://doi.org/10.1177/0093650219898094

Vraga, E. K., & Tully, M. (2021). News literacy, social media behaviors, and skepticism toward information on social media. *Information, Communication & Society*, *24*(2), 150–166. https://doi.org/10.1080/1369118X.2019.1637445

Vraga, E. K., Tully, M., & Bode, L. (2022). Assessing the relative merits of news literacy and corrections in responding to misinformation on Twitter. *New Media & Society*, *24*(10), 2354–2371. https://doi.org/10.1177/1461444821998691

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, *85*(3), 423–441. https://doi.org/10.1080/03637751.2018.1467564

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe report, *DGI (2017)*, *9*. Retrieved from: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, *9*(11), 39–52. https://doi.org/10.22215/timreview/1282

Xiao, X., Su, Y., & Lee, D. K. L. (2021). Who consumes new media content more wisely? Examining personality factors, SNS use, and new media literacy in the era of misinformation. *Social Media & Society*, *7*(1), 1–12. https://doi.org/10.1177/2056305121990635

Yang, Y., Davis, T., & Hindman, M. (2021). *Visual misinformation on Facebook: A work in progress paper*. Retrieved from: https://www.cmu.edu/ideas-social-cybersecurity/events/conference-archive/archive-conference-2021/conference-papers-2021/conference-2021-paper-8-visualmisinformation.pdf