



# User agency–based versus machine agency–based misinformation interventions: The effects of commenting and AI fact-checking labeling on attitudes toward the COVID-19 vaccination

new media & society

1–21

© The Author(s) 2023

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/14614448231163228

[journals.sagepub.com/home/nms](https://journals.sagepub.com/home/nms)



Jiyoung Lee  and Kim Bissell

The University of Alabama, USA

## Abstract

This study aimed to examine the effects of commenting on a Facebook misinformation post by comparing a user agency–based intervention and machine agency–based intervention in the form of artificial intelligence (AI) fact-checking labeling on attitudes toward the COVID-19 vaccination. We found that both interventions were effective at promoting positive attitudes toward vaccination compared to the misinformation-only condition. However, the intervention effects manifested differently depending on participants' residential locations, such that the commenting intervention emerged as a promising tool for suburban participants. The effectiveness of the AI fact-checking labeling intervention was pronounced for urban populations. Neither of the fact-checking interventions showed salient effects with the rural population. These findings suggest that although user agency- and machine agency–based interventions might have potential against misinformation, these interventions should be developed in a more sophisticated way to address the unequal effects among populations in different geographic locations.

## Keywords

AI fact-checking label, commenting, COVID-19 vaccine misinformation, online experiment, vaccine attitudes

---

## Corresponding author:

Jiyoung Lee, Department of Media and Communication, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Republic of Korea.

Email: [jiyoung.lee@g.skku.edu](mailto:jiyoung.lee@g.skku.edu)

Despite the medical consensus that the COVID-19 vaccine is a key to protecting oneself and one's community from the pandemic, vaccine hesitancy has been a persistent concern in the United States (Coustasse et al., 2021). Among other factors, misinformation and conspiracy theories have played a prominent role in increasing vaccine hesitancy. Conspiratorial misinformation has been spread particularly through social media.

Given the growing concerns over the impact of conspiratorial misinformation about the COVID-19 vaccine, a wide range of entities (e.g. social media platforms, journalists, and third-party fact-checking organizations) have adopted fact-checking practices. Researchers have taken efforts to test a variety of forms of fact-checking interventions, such as labeling misinformation (e.g. Zhang et al., 2021a) or nudging audiences to think about the accuracy of information (Pennycook et al., 2020). Moreover, recent fact-checking interventions have applied artificial intelligence (AI) to seek out automated ways to detect misinformation (e.g. Nakov et al., 2021), but research in this area is far from conclusive.

Building on previous literature, the current research takes a comprehensive approach to examining social media platform-based interventions by focusing on two misinformation interventions manipulated in terms of user/machine agency: (a) commenting on the misinformation post (i.e. a user agency-based misinformation intervention where users are given control over the intervention process) and (b) AI fact-checking labeling (i.e. machine agency-based misinformation intervention which offers sole agency to the [AI] system). We analyzed and compared the degree to which these interventions reduced the impact of conspiratorial misinformation on COVID-19 vaccine attitudes. In addition, we examined how our misinformation interventions elicited different effects among urban, suburban, and rural users. The sample for the present study was focused on Alabama residents, considering the large health disparities between urban and rural residents in the state. For example, Black Belt counties in Alabama are known to have limited healthcare providers and a high number of uninsured residents, which is directly linked to their limited access to healthcare (Barry-Jester, 2017). Particularly during the COVID-19 pandemic, Alabama has become one of the states that has experienced tremendous health disparities between rural and urban residents, such that residents in rural counties have been markedly susceptible to COVID-19 (Crozier et al., 2021). Addressing this additional question provided insights on the appropriateness of implementing the one-size-fits-all approach to misinformation interventions.

## **Social media misinformation and attitudes toward COVID-19 vaccination**

Misinformation about the COVID-19 vaccine has challenged the efforts to promote positive attitudes toward vaccination, amplifying concerns about the efficacy and safety of the COVID-19 vaccine among the public. Vulnerability to anti-vaccine misinformation has been attributed to various social media features, including a lack of professional journalistic gatekeepers and authoritative voices (Carlson, 2020) and/or active attempts of malicious actors (e.g. bots, trolls) to spread misinformation virally (Broniatowski et al., 2018). Especially when information supply is low and uncertainties are prevalent, users are likely to "turn to unofficial sources to satisfy their information needs," such as

their social media networks, neighbors, family, or other close-knit community members (Heverin and Zach, 2012: 35), which might, in turn, increase the likelihood of being exposed to misinformation.

While Vraga et al. (2022) found no significant effects of exposure to misinformation on misperceptions, exposure to biased information online, in general, contributes to inaccurate beliefs (Garrett et al., 2016). The effect of exposure to misinformation can linger even in the presence of a correction, creating what is called a belief echo (Thorson, 2016). Therefore, our first hypothesis tests whether exposure to a misinformation post about the COVID-19 vaccine without any misinformation interventions, compared to non-exposure, would result in more negative attitudes toward COVID-19 vaccination.

H1: Individuals who are exposed to a misinformation post only without any interventions (i.e. misinformation-only condition) will be less likely to hold positive attitudes toward the COVID-19 vaccination compared to those who are not exposed to a misinformation post (i.e. control condition).

## **Commenting as user agency–based misinformation intervention**

Previous research has revealed that misinformation is less likely to provide complex details to facilitate heuristic processing of the content (Osatuyi and Hughes, 2018). Given this phenomenon, it seems necessary to examine how to prevent audiences from easily being influenced by a misinformation post. One way to defend oneself against misinformation is to cognitively scrutinize the misleading content rather than processing it superficially and accepting it as is (Pennycook et al., 2020). Previous research shows that the act of expressing oneself on social media leads to greater cognitive elaboration (e.g. Oeldorf-Hirsch, 2018; Yoo et al., 2017), which means that individuals process the information deeply by linking to their existing knowledge or paying attention to it. Therefore, we suggest that expression via commenting activities can lead users to pause and think about the truthfulness of the misleading content. Commenting on a misinformation post can fall into a *user agency–based intervention* because it gives users control over communicating about the misinformation content—in other words, it gives users control over whether to simply indicate their agreements on the misinformation post or add specific thoughts in the comments section.

In fact, social media platforms not only function as sources of information but also motivate users to express themselves, allowing them to exercise user agency. The interactive, bidirectional features of social media communication facilitate expression (Halpern and Gibbs, 2013). Self-expression on social media additionally takes various forms—for example, leaving comments on social media posts, replying to others' opinions, or creating posts on one's newsfeed. Furthermore, the interactive features of social media enable one's expression to be displayed to the public simultaneously, including one's social networks (Gil De Zúñiga et al., 2014). Hence, by expressing one's thoughts online rather than passively consuming information, users may conceive of themselves as issue-involved participants who act, as opposed to fence-sitters (Gil De Zúñiga et al., 2014).

The potential of a user agency-based misinformation intervention is fundamentally linked to the “sender effect” which Pingree (2007) had suggested before the rise of social media. The sender effect posits that expression activities can facilitate self-reflection and extensive cognitive activity. For example, Pingree (2007: 447) noted, “expression, not reception, may be the first step toward better citizenship,” which suggests that expression can “motivate exposure, attention and elaboration of media messages, and the act of message composition is often much more effective at improving understanding than any act of reception could be.” According to this view, composing and releasing messages may thus have effects that make senders reflect on one’s own thoughts, which can further contribute to their understanding (Bem, 1967). The sender effect of the expression on cognitive elaboration has garnered empirical support from previous literature, particularly in the context of social media. For example, Nekmat (2012) showed that the expression of personal messages increased the extent of cognitive elaboration that individuals expended. More specifically, in the context of social media, Oeldorf-Hirsch (2018) found that active social media news engagement such as commenting on or replying to news stories had a positive impact on cognitive elaboration. In the context of health, previous research suggests that the expression of Zika-related information on social media could facilitate extensive cognitive activities, given that it could induce greater tendencies to perceive the health crisis as risk to oneself or one’s affiliated community, which, in turn, could lead to increased intentions to seek information and preventive behavioral intentions (Lee et al., 2020).

However, few studies have tested the effects of user agency-based misinformation interventions to deal with misinformation-inducing problems, particularly related to health issues such as the COVID-19 vaccination. Specifically, previous research lacked clarity on how commenting on misinformation posts would have a direct effect on attitudes toward the health crisis issue. Addressing these limitations in the current study directs us to posit the following hypothesis about the effect of leaving comments on a misinformation post on generating positive attitudes toward the COVID-19 vaccination. Based on the assumption that the user agency-based intervention can reduce the negative impact of misinformation because it can prevent individuals from falling for misinformation, H2 is advanced:

H2: Individuals who are exposed to a misinformation post only without any interventions (i.e. misinformation-only condition) will be less likely to hold positive attitudes toward the COVID-19 vaccination, compared to those who leave comments on a misinformation post (i.e. misinformation comments condition).

## **AI fact-checking label as machine agency-based misinformation intervention**

In contrast to commenting, AI fact-checking labeling is an intervention solely driven by system-generated rules. This intervention labels the misinformation as false based on an automatic AI system’s decision-making process. Social media platforms (e.g. Google, Facebook) have widely applied AI-based algorithms to automatically detect misinformation as part of the solution for countering the threat posed by the falsification of information. Although social media platforms have allowed individual social media users to

partially take a role in fact-checking practices at first, automated fact-checking tools which identify, verify, and respond to false claims through machine-learning AI systems have been increasingly used as a cost-efficient way to assess the credibility of social media posts (Graves, 2018). Fact-checking labels tend to use a simple, but explicit, tag-line message that indicates a piece of information as false (e.g. “this website contains false information”). For example, media literacy tool *NewsGuard* has aided Internet users to discern credible information from misleading or biased information by providing succinct credibility ratings of news and information websites through browser extensions. Facebook’s disputed labeling practices also represent the accelerated efforts of social media platforms to verify content. Given that the AI fact-checking label is governed by AI systems which grant AI controllability, such intervention can fall into a *machine agency-based misinformation intervention*.

As much empirical research has shown, testing the effectiveness of fact-checking labels has garnered much scholarly attention in recent years. Despite some evidence on the limited effect of fact-checking labels (e.g. Jennings and Stroud, 2021; Oeldorf-Hirsch et al., 2020), studies at large have concluded that the presence of fact-checking labels is effective at correcting misinformation. As found in Zhang et al. (2021a), compared to those who were exposed to misinformation, participants who were exposed to fact-checking labels had more favorable attitudes toward vaccines overall, and such effects of fact-checking labels were shown regardless of participants’ baseline vaccine skepticism. Related results were found in other empirical studies, such that the flagging of misinformation reduced Facebook users’ likelihood of sharing false news (Mena, 2020) and beliefs in false posts (Lee, 2022). Although not related to flagging, Bode and Vraga (2015) also found that when related articles correcting misinformation were attached to misleading Facebook posts, misperceptions were significantly reduced. The results of these studies suggest that flagging misinformation via simple fact-checking labels has been effective; in other words, the presence of fact-checking labels can be leveraged as an effective tool on social media platforms. The implied truth effect, coined by Pennycook et al. (2020), describes a phenomenon that tagging misinformation as false increases the perceived accuracy of untagged false posts. This effect has been empirically supported in a series of studies that untagged stories were perceived to be correct, irrespective of the truthfulness of the information (Pennycook et al., 2020). Therefore, when original misleading statements were not explicitly flagged as false, fact-checking effects are likely to be limited (Walter et al., 2020). This account echoes the effectiveness of explicitly tagging misinformation as false.

However, there remains a question as to whether fact-checking labels that specifically indicate the *use of AI* in fact-checking practices are effective at reducing the influence of misinformation, which warrants further investigation. Indeed, only few studies have explored the effects of fact-checking interventions attributed to AI. In Zhang et al.’s (2021a) study, using a fact-checking label provided by algorithms (i.e. “This post is falsified. Fact-checked by Deep Learning Algorithms”) was rated lower on source expertise relative to fact-checking labeling attributed to universities and health institutions. This suggests that different types of sources were found to be effective at mitigating the negative impact of vaccine misinformation. Additional work by Horne et al. (2020) demonstrated that under novel and emerging topics (i.e. COVID-19) where individuals might experience uncertainties, the AI fact-checking label was more effective at helping audiences identify the truth of

news articles. Furthermore, such effectiveness was pronounced especially when the AI intervention directly suggests the trustworthiness in the news source (i.e. “Our smart AI system indicates this is a trusted news source) or content (i.e. “Our smart AI system rates this article as accurate and reliable) by leading people to pay attention to those cognitive heuristic cues (Horne et al., 2020). However, a recent study that tested the fact-checking label attributing to AI’s decision-making rules found that the effectiveness was shown only when being transparent about the rationales of AI’s decision-making rules, suggesting the importance of implementing explainable AI (Liu, 2021).

Despite the potential limited effect of the simple AI fact-checking label, attributing a fact-checking label to AI can be effective, according to the machine heuristic. A machine heuristic refers to a mental shortcut because users believe machines are unbiased and objective entities (Sundar, 2008). As argued by Sundar and Kim (2019: 2), “When the perceived locus of our interaction is a machine, rather than another human being, . . . we automatically apply common stereotypes about machines, namely that they are mechanical, objective, ideologically unbiased and so on.” Because individuals tend to assess algorithmic judgment as superior to decisions made by a human alone (Dawes et al., 1989), the fact-checking label attributed to AI can be evaluated as objective, unbiased, and reliable. As such, the AI fact-checking label can be leveraged as an effective tool that reduces the impact of misinformation posts about the COVID-19 vaccine, thereby increasing positive attitudes toward COVID-19 vaccination. Following this logic, we hypothesize:

H3: Individuals who are exposed to a misinformation post only without any interventions (i.e. misinformation-only condition) will be less likely to hold positive attitudes toward COVID-19 vaccination, compared to those who see the AI fact-checking label when seeing a series of posts including misinformation (i.e. AI fact-checking label condition).

## **Accounting for differences between rural versus urban residents**

The last aim of this study was to examine whether there were any differences in the proposed effects of the interventions on attitudes toward COVID-19 vaccination depending on individuals’ residential areas (i.e. rural, suburban, and urban) in the United States. Exploring differences by users’ residential locations is particularly relevant to the COVID-19 vaccination issue, as research has demonstrated that residents in rural and urban areas in the United States tend to show a large COVID-19-related knowledge gap (Zhang et al., 2021b). Research has revealed that COVID-19 was transmitted at a higher rate in densely populated urban areas at the initial stage of the pandemic (Amram et al., 2020), but as the pandemic progressed, rural residents faced more risks of COVID-19 (Paul et al., 2020).

Several factors have exacerbated the risks among residents from rural counties, which include lack of accessibility to credible and comprehensive news media (termed as “news deserts”), lower education, lower income, and poorer health care access than urban and suburban residents (Abernathy, 2018; Henning-Smith, 2020). In fact, the rural–urban

disparity has appeared to be a general phenomenon globally. For example, previous research found that rural residents in China were more likely to have lower levels of information appraisal which reflects the abilities to utilize health information to protect one's health, and this in case negatively affected preventive health behaviors against COVID-19 (Chen and Chen, 2020). In addition, most rural residents in Vietnam were found to have access to technology devices but less than half of them used the devices for the Internet or mobile apps (Yun Lee et al., 2022). Similarly, in the United States, although rural residents have adopted digital technologies at an increasing rate over the past years, digital divides persist between rural, suburban, and urban Americans, continuing the lack of connectivity in rural communities (Vogels, 2021). Digital technology access does not guarantee equitable access to accurate information, which can close health literacy gaps during a public health crisis (Jackson et al., 2021).

Moreover, the persistent digital divide between rural and urban residents mirrors their existing inequalities in sociodemographic characteristics. Evidence suggests that the lack of economic and educational resources along with an increasingly aging population in rural communities is related to residents' lack of motivation and need to access information (Correa and Pavez, 2016). In addition, rural residents' tendencies to rely on word-of-mouth or certain media channels for COVID-19 information have led them to be more susceptible to misinformation that downplayed the risk of COVID-19 and affected their willingness to take the COVID-19 vaccine (Doughty and Fahs, 2021). Particularly, in southern states like Alabama where many rural locations have a larger proportion of underserved populations, disparities in COVID-19 health literacy are more pronounced (Mueller et al., 2021). This has been supported by recent research which found from a statewide survey that urban residents in Alabama were likely to understand information about COVID-19 more easily and have less difficulty accessing the information than their rural counterparts (Crozier et al., 2021).

Although the rural–urban distinction does not capture all sociodemographic differences, it adds a nuanced perspective on how misinformation interventions operate differently to enhance positive attitudes toward the COVID-19 vaccine, depending on individuals' residential locations. For example, rural residents' low baseline levels of awareness about the benefits of the COVID-19 vaccination may mitigate the effects of misinformation interventions in the current study. Given the increasing demand to design tailored misinformation intervention strategies to specific populations, examining differences among individuals from different residential locations can provide practical utility. Because no empirical studies have considered residential locations in understanding the misinformation intervention effects, we pose this as a research question:

*RQ1:* How will participants' residential areas (rural vs suburban vs urban) moderate the effects of the experimental intervention conditions (the comments condition and AI fact-checking label condition) on attitudes toward COVID-19 vaccination?

## Method

A one-way between-subjects online experiment comparing four conditions (leaving comments to a misinformation post [i.e. comments condition] vs AI fact-checking label on

Facebook posts [i.e. AI fact-checking label condition] vs exposure to a misinformation post only without misinformation interventions [i.e. misinformation-only condition] vs no misinformation post [i.e. no misinformation condition]). The data were collected after receiving approval from the Institutional Review Board. Upon agreeing to participate, participants first filled out screening questions. These questions included residential area (Alabama state only; rural, urban, and suburban areas via stratified sampling with the targeted percentage of participants residing in rural areas as 50%, urban and suburban as 25%, respectively<sup>1</sup>), age (over 18 years old), and whether participants were Facebook account holders. Participants who were eligible to participate in this study were then directed to answer a pretest questionnaire measuring preexisting beliefs in the misleading statement about Gates' plan to use the COVID-19 vaccine to plant microchips in people, general attitudes toward COVID-19 vaccines, COVID-19 vaccine skepticism, and COVID-19 vaccine uptake. Then, participants were randomly assigned to one of the four conditions and answered posttest questions about their attitudes toward COVID-19 vaccination. Last, participants answered demographic questions such as gender, education, race, and political orientation. At the end of the study, they were debriefed about the purpose of this study and the fact that the post about Gates was false. The average completion time of this study was 15 minutes.

### *Participants*

All participants for this study were living in Alabama between the recruitment window of February and March 2021 via Amazon Mechanical Turk (MTurk). Amazon MTurk is widely used to recruit diverse participants (e.g. Casler et al., 2013) and known to be a comparable method for recruiting rural populations to other conventional sample methods (e.g. Saunders et al., 2021). Although MTurk does not provide representative samples of Alabama residents in terms of sociodemographic characteristics, it is a popular source of convenience sample to recruit diversified populations. A total of 712 MTurk workers first accessed the study; however, 15 participants indicated that they did not reside in Alabama, 79 indicated they were not Facebook users, and 241 did not complete the study. This left the final sample of 377 ( $N=183$  rural population [48.5%],  $N=90$  urban population [23.9%],  $N=104$  suburban population [27.6%]). Of the final sample, 36.3% were female, 82.2% were white (with 14.3% Black or African American), and 54.9% were high school or some college graduates.<sup>2</sup> The mean age of the participants was 48.56 ( $SD=17.68$ ;  $min=18$ ,  $max=87$ ).

### *Stimuli*

Participants across the conditions were asked to consider the scenario that a series of Facebook posts were uploaded in one of the Facebook community groups they were a member of. The misinformation post which stated, "Did anyone hear that Bill Gates is using the COVID-19 vaccine to plant microchips in people?" was provided along with the picture of Gates. In the comments condition, participants were required to leave short comments about the misinformation post and were informed their comments would be shared with other community members of the Facebook groups. In the other conditions (i.e. AI fact-checking label, misinformation-only, and no misinformation), participants



were instructed to read the given Facebook posts. Except for the no misinformation condition, we included the four filler posts in addition to the misinformation post (randomly ordered) for all conditions. For the no misinformation condition, a total of five filler posts (including four filler posts used for other conditions; all randomly ordered) were shown to participants. A fictitious user's name was used for the misinformation post to prevent any confounding effects that might occur from sources. All participants were given a minimum of 30 seconds to read all the Facebook posts and allowed to move onto posttest questions.

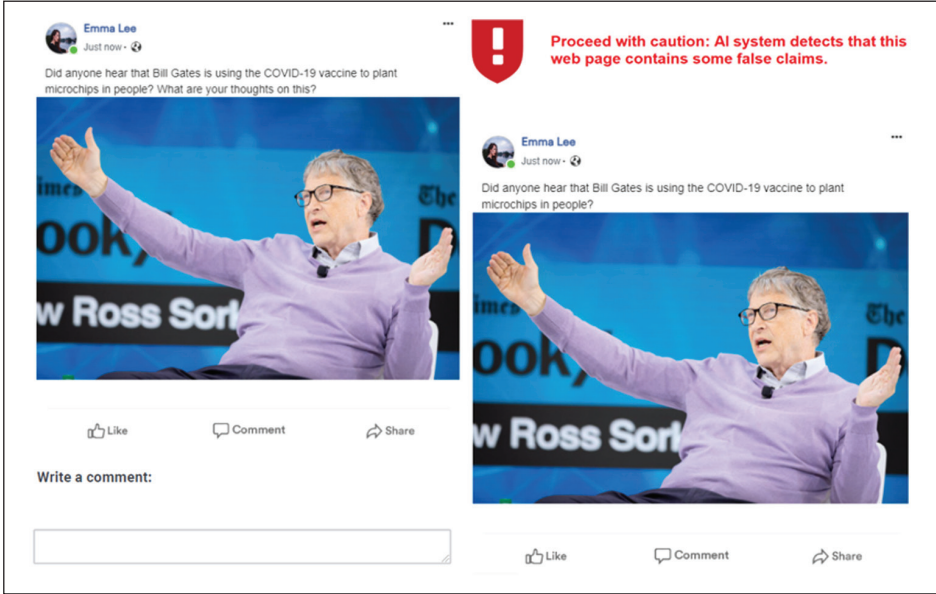
Specifically, for the comments condition, we manipulated the misinformation post that directly asked participants to share their thoughts in the comments section. For the AI fact-checking label condition, participants were first shown the AI label which reads, "Proceed with caution: an AI system detects that this web page contains some false claims." Then, they were directed to see a series of Facebook posts including the misinformation post. Although the fact-checking label we used for this study did not specifically indicate which misinformation post was found to be false, we followed the approach of other browser extensions' fact-checking labels. For example, NewsGuard, a tool that rates the credibility of online websites, uses a warning label (e.g. Proceed with Caution: This website severely violates basic journalistic standards) like the one used for this study. In this study, we modified a label to explicitly indicate that AI detected false claims. For the misinformation-only condition, we used the identical misinformation post used for the AI fact-checking label condition. For the no misinformation condition, the randomly ordered five given posts were irrelevant to the misinformation topic. In brief, participants in the other groups except for the comments condition were prevented from commenting (see Figure 1; see also Supplementary Materials for full stimuli including filler posts).

## Measurement

*(Posttest variable) Attitudes toward COVID-19 vaccination.* The dependent variable, attitudes toward COVID-19 vaccination, was measured in the posttest. Participants were asked to think about the following statement in the context of the having a child at the time of the study or if the participant were to have a child: "Getting the COVID-19 vaccine for my children would be. . ." (a) bad—good, (b) harmful—beneficial, (c) foolish—wise, (d) threatening—assuring, and (e) risky—safe. The items were adopted from previous research (Abhyankar et al., 2008) and measured on a 5-point bipolar scale. The items produced good reliability (Cronbach's  $\alpha = .97$ ,  $M = 3.41$ ,  $SD = 1.41$ ).

## Covariates

All variables used for covariates were measured before the stimuli (i.e. pretest variables) except political orientation. Political orientation was measured at the end of the study, along with other demographic variables. This approach was based on the possibility that covariates except for political orientation could have been affected by the stimuli, but demographic items like political orientation are individual traits, which are independent of the stimuli. In addition, measuring political orientation at the end might prevent priming effects in which participants could have assumed the study was related to their



**Figure 1.** Manipulation of the misinformation post for the comments condition (left), AI fact-checking label for the AI fact-checking label condition (upper right), and misinformation post for the AI fact-checking label condition and misinformation-only condition (lower right).

political ideology if presented at the beginning. Although participants were randomly assigned to one of the conditions, the following covariates were included in the model for the potential confounding effects in times of the pandemic.

**Preexisting beliefs in the misleading statement.** Given that the misinformation topic gained popularity at the time of collecting the data, we measured the extent to which participants held preexisting beliefs in the misleading statement. To prevent priming effects, five different false claims surrounding COVID-19 vaccines were shown to participants. Although we used five different claims, only the statement “Bill Gates is using the COVID-19 vaccine to plant microchips in people” was used as a covariate for our analysis. For each statement, participants were asked to indicate their believability levels measured on a 5-point scale (1 = *definitely false*, 5 = *definitely true*). The mean value of the Gates statement used for our analysis was 2.15 ( $SD = 1.34$ ).

**General attitudes toward COVID-19 vaccine.** For this covariate, participants answered a single item which asked them to indicate their general attitudes toward COVID-19 vaccines (1 = *very negative*, 5 = *very positive*;  $M = 3.29$ ,  $SD = 1.31$ ).

**COVID-19 vaccine skepticism.** For this covariate, we asked participants to rate how much they believed the following single statement as true or false: “The health risks of COVID-19 vaccines outweigh the benefits” (1 = *definitely false*, 4 = *definitely true*;  $M = 2.31$ ,  $SD = 1.02$ ). This item measuring skepticism was borrowed from Zhang et al. (2021a).

**COVID-19 vaccine uptake.** Considering that some participants might have been already vaccinated for COVID-19 at the time of data collection, we additionally asked whether participants were vaccinated for COVID-19 or not (0=No, 1=Yes). The percentage of participants who were not vaccinated was higher (79.8%) than that of participants who were vaccinated (20.2%).

**Political orientation.** As for the political orientation variable, we asked, “when it comes to politics, would you consider yourself . . .” (1=*very conservative*, 7=*very liberal*;  $M=3.25$ ,  $SD=1.76$ ).

## Results

Before running analyses for the hypotheses and research question, we conducted a correlation analysis among residential areas and measured covariates. The correlation results are presented in Table 1. To explain briefly, the results showed that urban residents of our sample were more likely to hold preexisting beliefs in misinformation, compared to rural residents ( $r=.19$ ,  $p<.01$ ), and suburban residents, compared to rural, were more likely to hold positive attitudes toward the vaccine ( $r=.18$ ,  $p<.01$ ) and less likely to hold vaccine skepticism ( $r=-.12$ ,  $p<.05$ ). Preexisting beliefs in the misinformation statement were negatively associated with positive attitudes toward the COVID-19 vaccine ( $r=-.39$ ,  $p<.001$ ) but positively associated with COVID-19 vaccine skepticism ( $r=.44$ ,  $p<.001$ ). Vaccine skepticism and positive attitudes toward the vaccine showed a negative association ( $r=-.46$ ,  $p<.001$ ). Regarding COVID-19 vaccine uptake, a positive association was shown with positive attitudes toward the vaccine ( $r=.38$ ,  $p<.001$ ), while a negative association was shown with COVID-19 vaccine skepticism ( $r=-.13$ ,  $p<.01$ ). As for political orientation, individuals with more liberalism were less likely to hold preexisting beliefs in misinformation ( $r=-.16$ ,  $p<.01$ ), more likely to hold positive attitudes toward the vaccine ( $r=.16$ ,  $p<.01$ ), and less likely to hold vaccine skepticism ( $r=-.12$ ,  $p<.05$ ).

To test H1, participants in the misinformation-only condition and no misinformation condition were selected. A one-way analysis of covariance (ANCOVA) treating these two conditions as predictors of attitudes toward COVID-19 vaccination was run with covariates. The result did not support H1 because there was no significant difference in attitudes toward COVID-19 vaccination between the misinformation-only condition and the no misinformation condition,  $F(1, 198)=.95$ ,  $p=.33$ ,  $\eta_p^2=.01$ . Although not hypothesized, the additional results on the main effect of the misinformation-only condition (vs no misinformation) in each group of residential areas revealed the non-significant effects across rural,  $F(1, 94)=2.53$ ,  $p=.12$ ,  $\eta_p^2=.03$ ; urban,  $F(1, 45)=0.82$ ,  $p=.37$ ,  $\eta_p^2=.02$ ; and suburban,  $F(1, 45)=1.46$ ,  $p=.23$ ,  $\eta_p^2=.03$ , groups.

H2 and H3 were tested by selecting participants assigned to one of the experimental conditions: the comments condition, the AI fact-checking label condition, or the misinformation-only condition. As expected, a significant difference in attitudes toward the COVID-19 vaccination was shown by the three experimental groups, according to a one-way ANCOVA,  $F(1, 266)=3.40$ ,  $p<.05$ ,  $\eta_p^2=.03$ . Post hoc tests with Fisher’s least significant differences (LSD) revealed that participants who left the comments on the misinformation showed a higher level of positive attitudes toward vaccination

**Table 1.** Pearson correlation results on residential areas and measured covariates.

Variable name	1	2	3	4	5	6	7
1. Rural (0) vs Urban (1)	–						
2. Rural (0) vs Suburban (1)	NA	–					
3. Preexisting beliefs ( $M=2.15$ , $SD=1.34$ ; 5-point scale)	.19**	-.10	–				
4. General attitudes toward COVID-19 vaccine ( $M=3.29$ , $SD=1.31$ ; 5-point scale)	.07	.18**	-.39***	–			
5. COVID-19 vaccine skepticism ( $M=2.31$ , $SD=1.02$ ; 4-point scale)	.12	-.12*	.44***	-.46***	–		
6. COVID-19 vaccine uptake (79.8% not vaccinated)	.08	.09	-.05	.38***	-.13*	–	
7. Political orientation (Liberalism; $M=3.25$ , $SD=1.76$ ; 7-point scale)	.06	-.003	-.16**	.16**	-.12*	-.08	–

NA: not applicable; SD: standard deviation.

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

( $M_{adjusted}=3.55$ ,  $SE=.10$ ), compared to those who saw the misinformation post only ( $M_{adjusted}=3.24$ ,  $SE=.10$ ),  $p < .05$ . Therefore, H2 was supported. Participants who saw the AI fact-checking label also held more positive attitudes toward vaccination ( $M_{adjusted}=3.58$ ,  $SE=.12$ ) than those who saw the misinformation post only, and this difference was significant ( $p < .05$ ). This lends support to H3. Although not listed as a hypothesis, there was no significant difference between the AI fact-checking label condition and the comments condition ( $p=.86$ ), suggesting that the interventions produced similar effects. In addition, when comparing each intervention to the control condition, no significant differences were found between the commenting and the control conditions ( $p=.31$ ) and the AI fact-checking label and the control conditions ( $p=.31$ ).

To examine whether the effects of experimental groups on attitudes toward COVID-19 vaccination are different among rural, suburban, and urban populations (RQ1), a two-way ANCOVA was performed with the experimental groups, residential areas as the independent factors, and attitudes toward COVID-19 as the dependent variable, with the inclusion of covariates (see Table 2). The results with all five covariates included revealed a statistically significant interaction effect,  $F(4, 260)=3.90$ ,  $p < .01$ ,  $\eta_p^2=.06$ , such that the main effects of the experimental conditions were moderated by residential areas.

Follow-up analyses of the simple main effect analysis using LSD comparisons showed that when leaving the comments on the misinformation post (i.e. the comments condition), rural populations showed significantly higher positive attitudes toward the COVID-19 vaccination ( $M_{adjusted}=3.57$ ,  $SE=.14$ ) than urban populations ( $M_{adjusted}=3.02$ ,  $SE=.21$ ), and this difference was significant,  $p < .01$ . The commenting activity was effective at increasing positive attitudes toward the vaccine for suburban populations as well ( $M_{adjusted}=3.98$ ,  $SE=.19$ ) relative to urban populations,  $p < .001$ . The difference between rural and suburban was also significant ( $p < .05$ ), such that higher positive attitudes toward the vaccination were found among suburban residents compared to rural residents when commenting on the misinformation post. Therefore, commenting as a user

**Table 2.** Two-way ANCOVA results on the interaction effects between experimental conditions and residential areas on attitudes toward COVID-19 vaccination.

	Type III sum of squares	df	Mean square	F	p	$\eta_p^2$
Corrected model	257.121 <sup>a</sup>	13	19.78	21.34	***	.52
Intercept	31.13	1	31.13	33.58	***	.11
Experimental conditions <sup>b</sup>	5.66	2	2.83	3.06	.05	.02
Residential areas <sup>c</sup>	1.47	2	0.74	0.79	.45	.01
Experimental conditions <sup>b</sup> × Residential areas <sup>c</sup>	14.46	4	3.61	3.90	.004	.06
Preexisting beliefs	6.06	1	6.06	6.54	.01	.03
General attitudes toward COVID-19 vaccine	69.28	1	69.28	74.75	***	.22
COVID-19 vaccine skepticism	1.02	1	1.02	1.10	.30	.004
COVID-19 vaccine uptake	10.99	1	10.99	11.86	.001	.04
Political orientation	3.74	1	3.74	4.03	.05	.02
Error	240.96	260	0.93			

N = 274.

<sup>a</sup> $R^2 = .52$  (adjusted  $R^2 = .49$ ),  $F(4, 260) = 3.90$ ,  $p < .01$ .

<sup>b</sup>Experimental conditions: comments, AI fact-checking label, and misinformation-only.

<sup>c</sup>Residential areas: rural, suburban, urban.

\*\*\* $p < .001$

agency-based intervention elicited the pronounced effect on increasing positive attitudes toward the COVID-19 vaccination among suburban residents.

For the AI fact-checking label condition, although the difference did not reach a statistically significant level ( $p = .12$ ), slightly higher positive attitudes toward vaccination were found among urban residents ( $M_{adjusted} = 4.03$ ,  $SE = .24$ ) relative to rural residents ( $M_{adjusted} = 3.53$ ,  $SE = .18$ ). The effect of the AI fact-checking label on positive attitudes toward the vaccination among urban residents was pronounced when compared to suburban residents ( $M_{adjusted} = 3.38$ ,  $SE = .19$ ;  $p < .05$ ). No difference was found between rural and suburban residents ( $p = .65$ ). Therefore, the AI fact-checking label turned out to be the most effective among urban populations while neither of the interventions (i.e. commenting and the AI fact-checking label) among rural residents was effective. Meanwhile, no differences in attitudes toward vaccination were found within the misinformation-only condition between rural ( $M_{adjusted} = 3.12$ ,  $SE = .14$ ) and urban residents ( $M_{adjusted} = 3.33$ ,  $SE = .19$ ;  $p = .36$ ), rural and suburban residents ( $M_{adjusted} = 3.39$ ,  $SE = .19$ ;  $p = .24$ ), and suburban and urban residents ( $p = .82$ ). Table 3 and Figure 2 present the moderating effects of residential location.

## Discussion

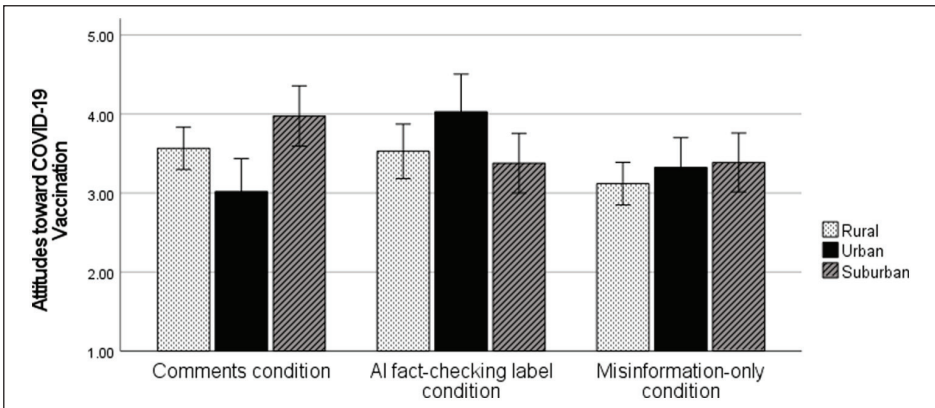
By responding to a persistent call to design effective misinformation interventions for mitigating misperceptions, the findings of the current study provide several important

**Table 3.** Descriptive statistics for attitudes toward COVID-19 vaccination by experimental conditions and residential areas.

	Comments condition (N=99)	AI fact-checking label condition (N=73)	Misinformation- only condition (N=102)
Residential areas	$M_{adjusted}$ (SE)	$M_{adjusted}$ (SE)	$M_{adjusted}$ (SE)
Rural (N=132)	3.57 (.14) <sup>a</sup>	3.53 (.18) <sup>de</sup>	3.12 (.14) <sup>f</sup>
Urban (N=64)	3.02 (.21) <sup>b</sup>	4.03 (.24) <sup>e</sup>	3.33 (.19) <sup>f</sup>
Suburban (N=78)	3.98 (.19) <sup>c</sup>	3.38 (.19) <sup>d</sup>	3.39 (.19) <sup>f</sup>

SE: standard error.

Mean values sharing identical superscripts within columns are not significantly different at the  $p < .05$  level. The results controlled for preexisting beliefs in the misleading statement, general attitudes toward COVID-19 vaccine, COVID-19 vaccine skepticism, COVID-19 vaccine uptake, and political orientation.



**Figure 2.** Comparison of adjusted mean scores for experimental conditions (misinformation-only, comments, and AI fact-checking label) by residential areas (rural, urban, suburban) on attitudes toward COVID-19 vaccination.

Higher values of attitudes toward COVID-19 vaccination indicate more positive attitudes toward the vaccination. The results controlled for preexisting beliefs in the misleading statement, general attitudes toward COVID-19 vaccine, COVID-19 vaccine skepticism, COVID-19 vaccine uptake, and political orientation.

implications in both theoretical and practical terms. As hypothesized, our social media platform-based misinformation interventions, manipulated in terms of user or machine-based agency, overall improved attitudes toward COVID-19 vaccination against misinformation. We therefore assert a user agency-based misinformation intervention holds promise, whereas users are led to engage in the misinformation post (supporting H2), and a machine agency-based misinformation intervention, which presents the AI

fact-checking label (supporting H3), in enhancing positive attitudes toward COVID-19 vaccination. This is meaningful, given that favorable attitudes can eventually reduce persistent concerns over the negative influences of health misinformation on users' pro-health attitudes.

To briefly elaborate on the results of H1, which compared the misinformation-only condition without any interventions to the control condition, this non-significant finding might be attributed to the possibility that some participants did not pay attention to the misinformation among the various posts presented; that said, participants who skimmed through the newsfeed might not even notice the misinformation post. However, it is worth noting from the additional analyses that the effects of misinformation interventions were salient when compared to the misinformation-only condition, given the non-significant differences between each intervention and the control condition when misinformation was not presented. This implies that activities like commenting or seeing the AI fact-checking label attached to the misinformation could be more effective at enhancing positive attitudes toward vaccines than simply seeing the misinformation without any of these interventions. Although examining whether misinformation interventions can directly motivate users to carefully think about the truthfulness of the misinformation post was beyond the scope of our study, we suggest that enhancing users' engagement in the misinformation post and/or employing the AI fact-checking label might serve as effective tools that could mitigate the impact of misinformation.

To explain implications of each intervention more specifically, our result on the effect of leaving comments on the misinformation post as a user agency-based intervention demonstrates the applicability of the "sender effect" (Pingree, 2007) as a theoretical framework to explain how users might come to hold attitudes toward vaccination when leaving their thoughts about the misinformation post. In line with the sender effect, which has demonstrated that expression yields cognitive benefits (Pingree, 2007; Shah, 2016), empirical research has found that the act of expression in the social media environment facilitates cognitive activities (e.g. Oeldorf-Hirsch, 2018) and motivates users to become active participants (Gil De Zúñiga et al., 2014).

Taking these previous findings into account, the current study expands the theoretical framework of the sender effect in the context of interactive social media by shedding light on the promising role of leaving comments on the misinformation post. This, in turn, might result in an increase in positive attitudes toward vaccination. Put differently, commenting on the misinformation post might serve to lead users to pause and rethink misleading content, instead of processing the conspiratorial claim heuristically. This type of a user agency-based interface, which allows users to actively engage in the misinformation post, would lead them to serve as additional sources of information. In this case, users would be more involved and empowered when processing the misinformation content. In addition to this explanation that commenting activities can enhance user agency in processing the misinformation post, users might know their comments would be simultaneously visible to the public—especially to their close social media networks—and this could have also influenced attitudes toward the vaccination against the conspiratorial claim. When leaving comments, users might recognize that they should be more careful in sharing their thoughts about the message content, perhaps because they might think they can serve as an additional source that directly impacts others' understanding on the vaccination issue.

We also found that the AI fact-checking label as a machine agency-based intervention (i.e. an intervention which offers sole agency to the [AI] system) was effective at promoting positive attitudes toward vaccination, which is seemingly in line with previous research on the effectiveness of the fact-checking label (e.g. Lee, 2022; Zhang et al., 2021a). However, our finding is unique from previous studies, given that our interventions applied the automated AI system to fact-checking practices. By explicitly attributing the fake post detection to the AI, the automated AI fact-checking label can produce a promising effect. This could suggest that users might perceive the AI as objective, neutral, and accurate, as posited by the machine heuristic (Sundar, 2008). Following this logic of the machine heuristic, the AI fact-checking label might have triggered users to be skeptical of social media posts they saw and resist or think more critically about the misinformation post. Perhaps because users might notice the falsehood of the misinformation post about Gates through the AI fact-checking label, users' positive attitudes toward vaccination might be shown in our finding. Therefore, even if the automated intervention does not offer agency to the individual user, employing the AI system can function as a cost-efficient tool.

However, we found different effects of misinformation interventions depending on where users reside (rural vs suburban vs urban). Specifically, the findings on the moderation effects of residential locations indicate that for rural respondents, the interventions did not generate salient effects on promoting positive attitudes toward COVID-19 vaccination. Although leaving comments on the misinformation post as a user agency-based intervention seemed to be effective, the commenting intervention elicited the most pronounced effect among suburban residents. For urban populations, it was the AI fact-checking label that emerged as an effective tool for promoting positive attitudes toward vaccination.

The salient effects of the AI fact-checking label for urban populations could be attributed to a gap between rural, suburban, and urban residents in understanding the operation of AI. Although not examined in this study, it could be possible that urban populations might have higher levels of understanding about the AI operation on fact-checking interventions. This finding also suggests the efforts to provide more detailed explanations on the AI mechanism of detecting misinformation, referred to as explainable AI (e.g. Miller et al., 2017; Rai, 2020). Explainable AI might prove beneficial for other populations such as rural residents who might need more knowledge on AI operation. Our argument may build upon recent research which suggests that the rationales for AI's fake news detection (i.e. being transparent about the AI's decision-making) should be revealed transparently to reduce uncertainty and increase trust (Liu, 2021). We further suggest that for people in rural/suburban locations who have lower media and health literacy levels than urban populations (e.g. Chen et al., 2019), providing the simple AI fact-checking label without adding the sufficient rationales of AI decision rules would not be sufficient to expect them to have positive attitudes toward the vaccine issue.

The less significant effect of the AI fact-checking label condition compared to the commenting intervention among rural/suburban populations suggests the need to employ explainable AI to the fact-checking label in layperson's terms to expect a significant intervention effect. In addition, there should be the need to develop more engaging interactive interface systems, given the potential of the commenting condition among rural/suburban



residents. Given the circumstances, the findings that our misinformation interventions on social media were not equally strong for all users echo the need to avoid implementing a “one-size-fits-all” approach to designing health misinformation interventions (Chou and Budenz, 2020). Although our participants were confined to Alabama residents, the less pronounced effects of the AI fact-checking label among rural and suburban can be applied to other contexts, given that the large gaps in health and digital media literacy disparities between rural and urban residents can be a common phenomenon across the world (e.g. Chen and Chen, 2020; Vogels, 2021). Therefore, future research should make continuous efforts to design effective, real-time fact-checking strategies programmed by AI to dissuade users, including rural residents, and counter misinformation to narrow the digital media literacy divide between rural and urban citizens.

This study has several limitations that could be addressed in future research. Given that Alabama has one of the lowest COVID-19 vaccination rates in the United States, we recruited only Alabama residents from the MTurk pool. However, we acknowledge that our sample does not represent the general population in the United States. The second limitation relates to the usage of a single misinformation topic, which could pose a challenge to generalizability. More research is warranted to investigate whether our findings can be replicated in different misinformation topics across health, science, and politics for various groups (e.g. media literacy levels, age, vaccination willingness). Future research is encouraged to measure other individual differences such as health literacy, AI literacy, preexisting attitudes toward the AI detection system, and levels of trust in social media platforms (Facebook herein), which have not been considered in this study. Related, while we are still able to infer from previous research that other demographic variables (e.g. education, income, and access to healthcare providers) may have an influence on health disparities in residential areas, we cannot conclude whether such individual differences may play determining roles in the effectiveness of misinformation interventions. Given these limitations, future research can investigate the underlying factors that may explain the limited effect of the AI fact-checking label among rural/suburban residents.

Another limitation relates to the commenting intervention for the user agency condition. Although forcing participants to leave comments ensured we could observe the effects of expressing one’s thoughts on the misinformation post via commenting, users in real social media environments can opt to leave comments on the posts. Hence, future research can adopt more realistic experimental treatments and examine whether there are any differences between users who choose to leave comments or not. The relatively high rate of drop-out responses and possibly low response quality among some participants should be also noted as a limitation of this study. Particularly for the user agency condition, we encourage future studies to examine whether there might be any different effects of the intervention depending on the quality of user responses.

Our findings offer practical implications for social media and AI practitioners who have tried to implement best practices for dealing with misinformation on the platforms. In the spread of conspiratorial claims that can directly plague scientific communities, it is imperative to understand that delivering simple messages that indicate the falsehood of misinformation may not be enough to promote accurate understanding of science-based health recommendations. Practitioners should not only seek out ways to direct

users to step back from any suspicious posts they may see on social media and engage in thoughtful cognitive processes. However, practitioners should also consider that a simple AI fact-checking label as a “black box” would not produce equal effects for every social media user. Given that the current study is among the first to include residential locations in examining the effects of misinformation interventions, future research on misinformation should focus more on individual differences and develop a strategy on when, how, and to whom various types of misinformation interventions on social media should be directed to affect users’ attitudes.

### Authors’ note

Jiyoung Lee is also affiliated to Sungkyunkwan University, Republic of Korea.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5C2A02088387) and the Quick Response Research Award Program supported by the National Science Foundation (NSF Award #1635593). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or Natural Hazards Center. This article was first published by the Natural Hazards Center Quick Response Research Award Program, with support from the National Science Foundation. Used with permission: <https://hazards.colorado.edu/research/quick-response-report/archives>.

### ORCID iD

Jiyoung Lee  <https://orcid.org/0000-0002-0800-9355>

### Supplemental material

Supplemental material for this article is available online.

### Notes

1. The original purpose of this study was to see differences between rural versus suburban/urban residents. However, we found no significant effects when merging suburban/urban residents. Therefore, the final analyses reported in this article set apart suburban and urban residents.
2. Additional analyses showed that the urban ( $M=3.67$ ,  $SD=1.64$ ;  $p<.01$ ) and suburban ( $M=4.12$ ,  $SD=1.52$ ,  $p<.001$ ) samples reported higher education than the rural sample ( $M=3.11$ ,  $SD=1.33$ ), respectively. Although not listed as a hypothesis, the correlation results presented that higher levels of education are associated with more favorable attitudes toward the COVID-19 vaccine ( $r=.25$ ,  $p<.001$ ), less vaccine skepticism ( $r=-.12$ ,  $p<.05$ ), and more likelihood of getting vaccinated ( $r=.22$ ,  $p<.001$ ). Education was measured on a 7-point scale ranging from 1 = *less than high school* to 7 = *doctorate degree*.

### References

- Abernathy PM (2018) *The Expanding News Desert*. Chapel Hill, NC: University of North Carolina Press. Available at: [https://www.cislm.org/wp-content/uploads/2018/10/The-Expanding-News-Desert-10\\_14-Web.pdf](https://www.cislm.org/wp-content/uploads/2018/10/The-Expanding-News-Desert-10_14-Web.pdf)

- Abhyankar P, O'connor DB and Lawton R (2008) The role of message framing in promoting MMR vaccination: evidence of a loss-frame advantage. *Psychology, Health and Medicine* 13(1): 1–16.
- Amram O, Amiri S, Lutz RB, et al. (2020) Development of a vulnerability index for diagnosis with the novel coronavirus, COVID-19, in Washington State, USA. *Health & Place* 64: 102377.
- Barry-Jester AM (2017) The health care system is leaving the southern Black Belt behind. *FiveThirtyEight*, 28 June. Available at: <https://fivethirtyeight.com/features/the-health-care-system-is-leaving-the-southern-black-belt-behind/>
- Bem DJ (1967) Self-perception: an alternative interpretation of cognitive dissonance phenomena. *Psychological Review* 74(3): 183–200.
- Bode L and Vraga EK (2015) In related news, that was wrong: the correction of misinformation through related stories functionality in social media. *Journal of Communication* 65(4): 619–638.
- Broniatowski DA, Jamison AM, Qi S, et al. (2018) Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health* 108(10): 1378–1384.
- Carlson M (2020) Fake news as an informational moral panic: the symbolic deviancy of social media during the 2016 US presidential election. *Information, Communication & Society* 23(3): 374–388.
- Casler K, Bickel L and Hackett E (2013) Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior* 29(6): 2156–2160.
- Chen X and Chen H (2020) Differences in preventive behaviors of COVID-19 between urban and rural residents: lessons learned from a cross-sectional study in China. *International Journal of Environmental Research and Public Health* 17(12): 4437.
- Chen X, Orom H, Hay JL, et al. (2019) Differences in rural and urban health information access and use. *The Journal of Rural Health* 35(3): 405–417.
- Chou WYS and Budenz A (2020) Considering emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. *Health Communication* 35(14): 1718–1722.
- Correa T and Pavez I (2016) Digital inclusion in rural areas: a qualitative exploration of challenges faced by people from isolated communities. *Journal of Computer-Mediated Communication* 21(3): 247–263.
- Coustasse A, Kimble C and Maxik K (2021) COVID-19 and vaccine hesitancy: a challenge the United States must overcome. *The Journal of Ambulatory Care Management* 44(1): 71–75.
- Crozier J, Christensen N, Li P, et al. (2021) Rural, underserved, and minority populations' perceptions of COVID-19 information, testing, and vaccination: report from a southern state. *Population Health Management*. Epub ahead of print 8 October. DOI: 10.1089/pop.2021.0216.
- Dawes RM, Faust D and Meehl PE (1989) Clinical versus actuarial judgment. *Science* 243(4899): 1668–1674.
- Doughty B and Fahs PS (2021) Why getting COVID-19 vaccines to rural Americans is harder than it looks. *The Conversation*, 4 January. Available at: <https://theconversation.com/getting-covid-19-vaccines-to-rural-americans-is-harder-than-it-looks-but-there-are-ways-to-lift-the-barriers-152377>
- Garrett RK, Weeks BE and Neo RL (2016) Driving a wedge between evidence and beliefs: how online ideological news exposure promotes political misperceptions. *Journal of Computer-Mediated Communication* 21(5): 331–348.
- Gil De Zúñiga H, Molyneux L and Zheng P (2014) Social media, political expression, and political participation: panel analysis of lagged and concurrent relationships. *Journal of Communication* 64(4): 612–634.

- Graves L (2018) Understanding the promise and limits of automated fact-checking. Available at: [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves\\_factsheet\\_180226%20FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf)
- Halpern D and Gibbs J (2013) Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior* 29(3): 1159–1168.
- Henning-Smith C (2020) The unique impact of COVID-19 on older adults in rural areas. *Journal of Aging & Social Policy* 32(4–5): 396–402.
- Heverin T and Zach L (2012) Use of microblogging for collective sense-making during violent crises: a study of three campus shootings. *Journal of the American Society for Information Science and Technology* 63(1): 34–47.
- Horne BD, Nevo D, Adali S, et al. (2020) Tailoring heuristics and timing AI interventions for supporting news veracity assessments. *Computers in Human Behavior Reports* 2: 100043.
- Jackson DN, Trivedi N and Baur C (2021) Re-prioritizing digital health and health literacy in *Healthy People 2030* to affect health equity. *Health Communication* 36(10): 1155–1162.
- Jennings J and Stroud NJ (2021) Asymmetric adjustment: partisanship and correcting misinformation on Facebook. *New Media & Society*. Epub ahead of print 28 June. DOI: 10.1177/14614448211021720.
- Lee J (2022) The effect of web add-on correction and narrative correction on belief in misinformation depending on motivations for using social media. *Behaviour & Information Technology* 41(3): 629–643.
- Lee J, Kim JW and Chock TM (2020) From risk butterflies to citizens engaged in risk prevention in the Zika virus crisis: focusing on personal, societal and global risk perceptions. *Journal of Health Communication* 25(9): 671–680.
- Liu B (2021) In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human–AI interaction. *Journal of Computer-Mediated Communication* 26(6): 384–402.
- Mena P (2020) Cleaning up social media: the effect of warning labels on likelihood of sharing false news on Facebook. *Policy & Internet* 12(2): 165–183.
- Miller T, Howe P and Sonenberg L (2017) Explainable AI: beware of inmates running the asylum. *IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI)*. Available at: <https://arxiv.org/abs/1712.00547>
- Mueller JT, McConnell K, Burow PB, et al. (2021) Impacts of the COVID-19 pandemic on rural America. *Proceedings of the National Academy of Sciences* 118(1): 2019378118.
- Nakov P, Corney D, Hasanain M, et al. (2021) Automated fact-checking for assisting human fact-checkers. *arXiv preprint arXiv:2103.07769*.
- Nekmat E (2012) Message expression effects in online social communication. *Journal of Broadcasting & Electronic Media* 56(2): 203–224.
- Oeldorf-Hirsch A (2018) The role of engagement in learning from active and incidental news exposure on social media. *Mass Communication and Society* 21(2): 225–247.
- Oeldorf-Hirsch A, Schmierbach M, Appelman A, et al. (2020) The ineffectiveness of fact-checking labels on news memes and articles. *Mass Communication and Society* 23(5): 682–704.
- Osatuyi B and Hughes J (2018) A tale of two internet news platforms—real vs. fake: an elaboration likelihood model perspective. In: *Proceedings of the 51st Hawaii international conference on system sciences*, 3 January, pp. 3986–3994. Available at: <https://scholarspace.manoa.hawaii.edu/items/6f416a4e-349c-40f3-8c91-0d5631c717e6>
- Paul R, Arif AA, Adeyemi O, et al. (2020) Progression of COVID-19 from urban to rural areas in the United States: a spatiotemporal analysis of prevalence rates. *The Journal of Rural Health* 36(4): 591–601.

- Pennycook G, McPhetres J, Zhang Y, et al. (2020) Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science* 31(7): 770–780.
- Pingree RJ (2007) How messages affect their senders: a more general model of message effects and implications for deliberation. *Communication Theory* 17(4): 439–461.
- Rai A (2020) Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science* 48(1): 137–141.
- Saunders EC, Budney AJ, Cavazos-Rehg P, et al. (2021) Comparing the feasibility of four web-based recruitment strategies to evaluate the treatment preferences of rural and urban adults who misuse non-prescribed opioids. *Preventive Medicine* 152: 106783.
- Shah DV (2016) Conversation is the soul of democracy: expression effects, communication mediation, and digital media. *Communication and the Public* 1(1): 12–18.
- Sundar S (2008) The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Metzger MJ and Flanagin J (eds) *Digital Media, Youth, and Credibility*. Cambridge: MIT Press, pp. 72–100.
- Sundar SS and Kim J (2019) Machine heuristic: when we trust computers more than humans with our personal information. In: *Proceedings of the 2019 CHI conference on human factors in computing systems*, Glasgow, 4–9 May, pp. 1–9. New York: ACM.
- Thorson E (2016) Belief echoes: the persistent effects of corrected misinformation. *Political Communication* 33(3): 460–480.
- Vogels EA (2021) Some digital divides persist between rural, urban and suburban America. *Pew Research Center*, 19 August. Available at: <https://www.pewresearch.org/fact-tank/2021/08/19/some-digital-divides-persist-between-rural-urban-and-suburban-america/>
- Vraga E, Tully M and 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.
- Walter N, Cohen J, Holbert RL, et al. (2020) Fact-checking: a meta-analysis of what works and for whom. *Political Communication* 37(3): 350–375.
- Yoo SW, Kim JW, Gil De and Zúñiga H (2017) Cognitive benefits for senders: antecedents and effects of political expression on social media. *Journalism & Mass Communication Quarterly* 94(1): 17–37.
- Yun Lee H, Wang K, Deavours DF, et al. (2022) Information and communication technology use for seeking health information in Vietnamese living in rural areas. *International Journal of Health Promotion and Education*. Epub ahead of print 26 April. DOI: 10.1080/14635240.2022.2059694.
- Zhang J, Featherstone JD, Calabrese C, et al. (2021a) Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. *Preventive Medicine* 145: 106408.
- Zhang J, Zhu L, Li S, et al. (2021b) Rural–urban disparities in knowledge, behaviors, and mental health during COVID-19 pandemic: a community-based cross-sectional survey. *Medicine* 100(13): e25207.

## Author biographies

Jiyoung Lee is an Assistant Professor in the Department of Media and Communication at Sungkyunkwan University. Her research interests lie in emerging media effects (e.g. AI, AR) on the (mis)information flow in health and risk contexts. Her recent publications have been related to mechanisms of the persuasive effects of various types of misinformation including deepfake and developing media literacy interventions.

Kim Bissell serves as the Director for the Institute for Communication and Information Research and the co-director of the Health Communication Lab. Her research interests lie in the intersection of media, health, and children, and her recent publications have been related to the role of media in the development of children’s nutritional knowledge and attitudes.