



Fighting fake news on social media: a comparative evaluation of digital literacy interventions

Anat Toder Alon¹ · Ilan Daniels Rahimi² · Hila Tahar^{1,3}

Accepted: 13 January 2024 / Published online: 5 February 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Effective digital literacy interventions can positively influence social media users' ability to identify fake news content. This research aimed to (a) introduce a new experiential training digital literacy intervention strategy, (b) evaluate the effect of different digital literacy interventions (i.e., priming critical thinking and an experiential training exercise) on the perceived accuracy of fake news and individuals' subsequent online behavioral intentions, and (c) explore the underlying mechanisms that link various digital literacy interventions with the perceived accuracy of fake news and online behavioral intentions. The authors conducted a study, leveraging online experimental data from 609 participants. Participants were randomly assigned to different digital literacy interventions. Next, participants were shown a Tweeter tweet containing fake news story about the housing crisis and asked to evaluate the tweet in terms of its accuracy and self-report their intentions to engage in online activities related to it. They also reported their perceptions of skepticism and content diagnosticity. Both interventions were more effective than a control condition in improving participants' ability to identify fake news messages. The findings suggest that the digital literacy interventions are associated with intentions to engage in online activities through a serial mediation model with three mediators, namely, skepticism, perceived accuracy and content diagnosticity. The results point to a need for broader application of experiential interventions on social media platforms to promote news consumers' ability to identify fake news content.

Keywords Digital literacy · Experiential intervention · Fake news · Skepticism · Perceived diagnosticity

Introduction

We live in a world where truth and trust are often tenuous, and skepticism and doubt are rampant. The COVID pandemic accelerated the erosion of trust worldwide, leading to an epidemic of misinformation combined with a failed trust system unable to confront the vast amount of available

information (Edelman, 2021). Fake news has become a phenomenon of major significance in internet-based media and is currently considered one of the greatest threats to democracy, journalism, and freedom of expression (Zhou & Zafarani, 2020). Recently, deepfake techniques, which can manipulate images or videos through deep learning technology, have been used to create fake news images or videos, significantly increasing social concerns (Kim et al., 2021). Indeed, powerful new technology simplifies the manipulation and fabrication of content, and social networks dramatically amplify falsehoods as they are shared by uncritical publics (Ireton & Posetti, 2018; Jungherr & Schroeder, 2021).

The battle against fake news on social media presents multiple challenges (Lutzke et al., 2019). First, social media sites are popular sources of news. According to a Pew Research Center survey, 50% of U.S. adults reported receiving news from social media “often” or “sometimes,” with nearly a third of Americans regularly receiving news on Facebook, a quarter on YouTube, and smaller shares on

✉ Anat Toder Alon
anat.toder@gmail.com

Ilan Daniels Rahimi
irahimi@ono.ac.il

Hila Tahar
TAHAR@bgu.ac.il

¹ School of Business Administration, Peres Academic Center, 10 Shimon Peres St., 7610202 Rehovot, Israel

² School of Business Administration, Ono Academic College, 104 Zahal St., 5500 Kiriat Ono, Israel

³ Department of Management, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Twitter (14%), Instagram (13%), TikTok (10%), and Reddit (8%) (Liedke & Matsa, 2022). Second, social media providers typically do not control the accuracy or sources of content posted on their platforms. The absence of effective filters and control mechanisms implies that uncertainty regarding providers' credibility and content accuracy is inherent in online posts (Jungherr & Schroeder, 2021). Finally, social media's rapid dissemination poses an especially critical challenge regarding false information because falsehoods are diffused significantly farther, faster, deeper, and more broadly than accurate information in all categories (Vosoughi et al., 2018).

This study aims to provide valuable insights into the understanding of fake news consumption and the impact of diverse digital literacy interventions. In response to Guess et al.'s (2020) call for comprehensive evaluations of diverse digital literacy interventions, emphasizing the need to explore more intensive digital literacy training models, we introduce a new experiential training digital literacy intervention strategy. We aim to investigate the efficacy of two interventions—priming critical thinking and an experiential training exercise—in improving individuals' capacity to assess the accuracy of online content. Additionally, we heed Guess et al.'s (2020) recommendation to investigate the mechanisms of digital literacy interventions. Consistent with this suggestion, our research makes a crucial contribution to the field of fake news by uncovering the underlying mechanisms through which various digital literacy interventions operate.

Literature review

The concept of digital media literacy captures the skills and strategies needed to successfully navigate a fragmented and complex information ecosystem (Eshet-Alkalai, 2004). Most people find it difficult to accurately judge the quality of information they encounter online because they lack the skills and contextual knowledge necessary to discriminate between high- and low-quality news items (Guess et al., 2020). Scholars have investigated various interventions aimed at improving the digital media literacy of news consumers (for a review, see Kozyreva et al., 2022). These interventions can generally be categorized as nudges, which target behaviors; boosts, which target competencies; and refutation strategies, which target beliefs.

Types of digital literacy interventions: the nudging approach

Several experimental studies have demonstrated that interventions that facilitate deliberation on social media platforms reduce intentions to share fake news (Pennycook &

Rand, 2021). Accuracy prompts, used to shift people's attention to the concept of accuracy, comprise one type of nudging intervention. Research has shown that bringing an accuracy motive to mind by asking users if a randomly selected headline is accurate reduces intentions to share fake news (Pennycook, McPhetres et al., 2020, 2021). This approach has also been successfully deployed in a large-scale field experiment on Twitter that showed that a single accuracy message made users more discerning in their subsequent sharing decisions (Pennycook et al., 2021). Another nudging intervention introduced friction (i.e., pausing to think) into a decision-making process to slow information sharing. In an online experiment, participants who paused to explain why a headline was true or false were less likely to share false information than control participants (Fazio, 2020). A notable advantage of the nudging approach is its ease of implementation on social media platforms, making it highly scalable (Pennycook, McPhetres et al., 2020; Pennycook & Rand, 2021).

Types of digital literacy interventions: the boosting approach

Another potential intervention for reducing belief in fake news involves equipping people with the skills needed to identify it. Lutzke et al. (2019) tested the effect of two simple interventions, both of which primed critical thinking, on individuals' evaluation of the credibility of real and fake news on Facebook regarding climate change. Participants either only read a series of guidelines for evaluating news online or read and then rated the importance of each guideline. The guidelines included questions about recognizing the news source, believability, professionalism, and political motivation. The authors provide experimental evidence that asking oneself these questions can significantly reduce trust, liking, and intending to share fake news. Similarly, Guess et al. (2020) evaluated a priming critical thinking intervention by exploring the effects of exposure to Facebook's "Tips to Spot False News." This intervention provided simple rules to help individuals evaluate the credibility of sources and identify indicators of problematic content without expending significant time or effort. Using data from preregistered survey experiments conducted around elections in the United States and India, Guess et al.'s (2020) results indicate that exposure to this intervention reduced the perceived accuracy of both mainstream and fake news headlines, but the effects on the latter were significantly greater. Another example of a boosting intervention designed to encourage fact-checking competencies is lateral reading. Lateral reading is a simple heuristic for online fact-checking that asks media users to conduct additional independent online searches to verify the trustworthiness of a claim or source. Such interventions can be applied in school curricula (Brodsky et al., 2021)

or via a simple pop-up that shows people how to practice lateral reading (Panizza et al., 2022). Indeed, research has shown that *students' evaluation of online sources improved significantly after a series of course-embedded activities* (Brodsky et al., 2021). In summary, an advantage of the boosting approach is the active engagement of individuals in the learning process, leading to more effective skill development and a deeper understanding of digital competencies. A key feature of these programs is their reliance on transparency, as they hinge on the active involvement of individuals (Kozyreva et al., 2022).

Types of digital literacy interventions: the refutation approach

An alternative class of interventions involves preemptive ('prebunking') and reactive ('debunking') refutation strategies (see Ecker et al., 2022). Refutations reduce false beliefs by providing factual information alongside explanations of why a piece of misinformation is false or misleading. Thus, their primary target is belief calibration (Lewandowsky et al., 2020). An example of this type of intervention is the inoculation approach, which is based on the hypothesis that it is possible to confer psychological resistance to deceitful persuasion attempts by exposing people to weakened versions of misinformation. Inoculation interventions have been shown to reduce the persuasiveness of misinformation by warning of misinformation and correcting specific false claims or identifying tactics used to promote it (e.g., Cook et al., 2017; van der Linden et al., 2017; Roozenbeek et al., 2022). This approach is a refutation strategy and a boost because it also builds competence in identifying misleading argumentation techniques.

A series of studies investigated this technique in the context of a recently developed fake news game, the Bad News Game (Basol et al., 2020; Modirrousta-Galian & Higham, 2023; Roozenbeek & van der Linden, 2019). In this game, participants are exposed to "weakened doses" of misinformation techniques to improve their ability to spot such techniques and thus "inoculate" them against misinformation. These studies found a significant reduction in the perceived accuracy of fake news presented in the form of Twitter tweets. Other studies have evaluated the effectiveness of interventions that involve attaching warnings to headlines of news stories that have been disputed by third-party fact-checkers (Pennycook, Bear et al., 2020), introducing credibility label (Duncan, 2022) or using truth rating scales (Amazeen et al., 2018). A noteworthy drawback of refutation approaches is their reliance on opt-in participation, meaning individuals must actively choose to engage with the inoculation technique. This poses a significant challenge, especially considering that those who would benefit most from misinformation inoculation, such as individuals

with lower cognitive reflection, may be less inclined to initiate and participate in extended inoculation processes (Pennycook & Rand, 2021).

Types of digital literacy interventions: a summary

The various digital literacy intervention strategies discussed above offer promising means to combat the spread of fake news. Recently, scholars have pointed to the need to compare the effects of different digital literacy interventions (Guess et al., 2020; Kozyreva et al., 2022). The primary objective of our research was to align with these recommendations by investigating the impact of two boost-based interventions. One of these interventions entailed providing tips geared towards enhancing critical thinking, while the other utilized an experiential training exercise. Our study aimed to assess how these interventions affected individuals' perceptions of fake news accuracy, their subsequent online behavioral intentions, and the underlying mechanisms through which these interventions operated.

The role of users' pre-dispositional skepticism

Skepticism can be described as a disposition towards cautious doubt, encouraging individuals to approach information critically by recognizing that it may be influenced by the perspective and motivations of its source (Feuerstein, 1999). It promotes reflective thought and careful evaluation of beliefs, fostering an attitude of questioning and critical thinking (Feuerstein, 1999). In particular, the construct of electronic word-of-mouth (eWOM) skepticism was established to describe Internet users' pre-dispositional suspicion and distrust toward all eWOM communications (X. J. Zhang et al., 2016). It relates to Internet users' concerns regarding the truthfulness of eWOM messages in general, senders' motivations, and identity (X. J. Zhang et al., 2016).

Research has demonstrated that skepticism towards information shared on social media is negatively related to the inclination to embrace conspiracy theories related to COVID-19 (Ahadzadeh et al., 2023). This implies that maintaining a skeptical stance towards content on social media serves as a protective factor, mitigating susceptibility to the effects of COVID-19 conspiracy theories (Ahadzadeh et al., 2023). This finding supports the notion that skepticism plays a negative predictive role in the acceptance of misinformation and disinformation by motivating individuals to question the sources of information (Lewandowsky et al., 2012) and fostering analytical thinking (Feuerstein, 1999), leading to a decrease in belief in fake news (Swami et al., 2014).

In our current study, we delve into the intricate relationship between users' general skepticism towards online news and their perceptions of the accuracy of fake news when various digital interventions are implemented. We aim to

uncover how interventions designed to enhance critical thinking and improve digital literacy may influence individuals' inclination toward suspicion and distrust of online news in general, ultimately affecting their evaluations of the accuracy of information in the digital realm. By examining the interplay between intervention strategies, skepticism, and the evaluation of fake news accuracy, our research seeks to contribute valuable insights into the effectiveness of interventions aimed at addressing the challenges posed by misinformation and promoting media literacy in the digital age.

The role of perceived content diagnosticity

Social media serves as a platform for people to stay informed about news and politics (Amsalem & Zoizner, 2023). However, it takes time, effort, and energy to determine the accuracy and relevance of news from the vast amount of information provided through social media (L. Zhang et al., 2016). Duffy and Ling (2020) note that individuals are more likely to share news when they believe it will be relevant to their intended audience and when it carries high perceived informational value. In accordance with relevance theory, human communication is guided by expectations of relevance (Sperber & Wilson, 1995). Building on Altay et al.'s (2022) concept of message interestingness, we consider message relevance to be akin to diagnosticity, reflecting its potential to trigger profound cognitive effects, such as generating rich inferences and significantly altering beliefs or intentions. Thus, perceived content diagnosticity refers to the extent to which a piece of information can distinguish between various interpretations of an issue and its potential solutions (Li et al., 2013).

Much research has demonstrated that helpful user-generated content (UGC) posted on social media platforms is crucial in reducing information overload and driving voting behavior and brand choice (Bigne et al., 2021; Zhuang et al., 2023). The current study maintains that the perceived content diagnosticity of a news message adds an essential dimension to our understanding of the spread of fake news content. Following Jiang and Benbasat's (2004) conceptualization, the present study maintains that a news message is deemed diagnostic if it provides information individuals perceive as helping them become familiar with, understand, and evaluate a subject. The current study investigates the extent to which individuals' online behavioral intentions are influenced by their perception of content diagnosticity when various digital interventions are applied. We aim to shed light on how the perceived value of news content in aiding individuals' familiarity with, understanding of, and ability to evaluate a subject impacts their online actions. By exploring this relationship in the context of digital interventions, we seek to deepen our comprehension of how content

diagnosticity plays a pivotal role in the dissemination and reception of fake news in the digital landscape.

Hypothesis development

The main premise of the present research was to evaluate the effectiveness of employing various digital literacy interventions in decreasing respondents' belief in false news stories and intentions to engage in online activities related to this fake news content (i.e., linking to further reading, sharing the post with other people, and commenting).

Impact of priming critical thinking intervention

A core component of research on the cognitive science of reasoning focuses on dual-process theories, according to which analytic thinking can override automatic, intuitive responses (Pennycook et al., 2021). Across numerous recent studies, the evidence supports the prediction that people who deliberate more will be less likely to believe false news content (Bago et al., 2020; Pennycook & Rand, 2019). In line with previous studies (Guess et al., 2020; Lutzke et al., 2019), we hypothesized the following:

H1a: Compared to a control group, exposure to a priming critical thinking intervention reduces belief in fake news stories. In other words, exposure to a priming critical thinking intervention decreases the perceived accuracy of fake news stories.

Studies have demonstrated that even people's ability to detect fake news by systematically judging it as less accurate than real news does not stop them from disseminating it (Pennycook, McPhetres et al., 2020). One hypothesis is that people share inaccurate news because they fail to think analytically about accuracy when deciding what to share (Pennycook et al., 2021). Indeed, research has shown that people are more likely to consider a news item's accuracy when deciding whether or not to share it if they have simply been asked to do so (Fazio, 2020; Pennycook et al., 2021). These findings suggest that people can distinguish accurate from inaccurate news but that, unless specifically prompted, largely fail to use this ability in their sharing decisions (Pennycook & Rand, 2019). Consistent with previous research that revealed that an intervention that primed critical thinking (Lutzke et al., 2019) reduced individuals' sharing of fake news, we hypothesized that exposure to the tip intervention would reduce people's intentions to engage in online activities related to news content compared to a non-intervention condition:

H1b: Exposure to a priming critical thinking intervention reduces intentions to engage in online activities (i.e., comment, link, and share) related to fake news stories.

Impact of experiential training intervention

An experiential training intervention, by design, aims to encourage participants to practice distinguishing between fake and real content, thus promoting a spirit of inquiry. Kolb's theory of experiential learning (Kolb, 2014) views learning as a process of creating knowledge through the transformation of direct sensory experience and in-context action. Knowledge becomes the result of experiential understanding and transformation. Consequently, experiential learning is a process that enables learners to transform and create knowledge, skills, attitudes, and critical thinking. An experiential training intervention, falls under the boosting approach category of digital literacy interventions. This approach aims to equip individuals with practical skills and experiences that enable them to critically evaluate information, in this case, to distinguish between fake and real content. Experiential learning, encourages active engagement and hands-on experiences. By providing participants with opportunities to practice discerning between fake and real content, this intervention aligns with the goal of boosting digital competencies and improving media literacy. There is a paucity of literature that evaluates whether and how experiential intervention programs may be applied in the context of digital literacy. In the current research, we hypothesized that since the experiential interventions involve a hands-on approach, participants remain actively engaged throughout the training and are thus likely to internalize it. Therefore, the current study emphasized the advantages of experiential training intervention programs in enhancing media users' ability to identify fake news. Hence, we hypothesized the following:

H2a: Compared to a control group, exposure to an experiential training intervention reduces belief in fake news stories. In other words, exposure to an experiential training intervention decreases the perceived accuracy of fake news stories.

In addition, consistent with research showing that boost-based interventions (i.e., tips; Lutzke et al., 2019) reduce individuals' sharing of fake news, we hypothesized that exposure to the experiential training intervention would reduce people's intentions to engage in online activities related to news content compared to a non-intervention condition:

H2b: Exposure to an experiential training intervention reduces intentions to engage in online activities (i.e., comment, link, and share) related to fake news stories.

Users' skepticism and perceived content diagnosticity as mediators

This research aimed to shed light on the mechanisms through which different literacy interventions operate. Thus, the following hypotheses refer to the mediated, indirect effects of digital literacy interventions on perceived accuracy and intentions to engage in online activities related to news content through our two hypothesized mediators: news users' skepticism and perceived content diagnosticity.

Role of users' skepticism as a mediator

The presumed influences of digital literacy interventions on the perceived accuracy of fake news may be partially mitigated by users' skepticism, which may lower the level of belief in fake news stories. Accordingly, digital literacy interventions not only empower individuals to critically engage with online content but also serve as a foundational mechanism for the cultivation of users' pre-dispositional skepticism toward online news in general. This skepticism, in turn, acts as a potent tool in identifying and rejecting fake news stories (Ahadzadeh et al., 2023). Hence, we sought to understand whether literacy interventions influence perceptions of accuracy directly or indirectly through the elicitation of skepticism. If skepticism levels serve as a mediator, it would be helpful to know whether the underlying mechanism that links digital literacy interventions with perceived accuracy varies depending on the type of digital literacy intervention. In the case of fake news content, the rationale is as follows: the more news consumers perceive the motivations of news senders as ulterior and concealed, and the more they develop a distrust of online content, the better they will be able to identify fake news content. Thus, we formulated the following hypotheses:

H3a: Compared to a control group, exposure to (1) a priming critical thinking intervention or (2) an experiential training intervention increases users' skepticism.

H3b: Users' skepticism is negatively associated with perceived accuracy of fake news stories.

H3c: The difference in perceived accuracy between the control group and (1) a priming critical thinking intervention or (2) an experiential training intervention is mediated through users' skepticism.

Role of perceived content diagnosticity as a mediator

Following research on the significance of online customer review “diagnosticity” as the primary measure of how consumers evaluate a review (Li et al., 2013; Mudambi & Schuff, 2010), the current research assumed that the association between perceived accuracy of news items and intentions to engage in online behavior could be mediated through perceived content diagnosticity. Perceived content diagnosticity reflects the extent to which individuals perceive the content as informative, capable of generating rich inferences, and aiding in their understanding and evaluation of the subject matter. Hence, when people believe that the content of a news message is accurate, the message will be perceived as more diagnostic. Moreover, perceived diagnosticity is positively associated with users’ intentions to engage in online activities. In other words, if individuals perceive a news message as helping them to become familiar with, understand, and evaluate a subject, their intentions to engage in online activities related to its content will increase. Furthermore, we hypothesized that the association between perceived accuracy and users’ intentions to engage in online activities would be mediated by perceived content diagnosticity. Hence, we formulated the following hypotheses:

H4a: Perceived accuracy of fake news stories is positively associated with perceived content diagnosticity.

H4b: Perceived content diagnosticity of fake news stories is positively associated with intentions to engage in online activities (i.e., comment, link, and share) related to fake news stories.

H4c: The relationship between perceived accuracy and users’ intentions to engage in online activities is mediated through perceived content diagnosticity.

In summary, a theoretical serial mediation model was developed to examine the proposed relationship. We hypothesized that the relationship between media literacy interventions and users’ intentions to engage in online activities is serially mediated by users’ skepticism, perceived accuracy and perceived content diagnosticity. Specifically, media literacy interventions are hypothesized to enhance users’ predispositional skepticism toward online news by equipping them with critical thinking skills. Following skepticism, the next mediator in the chain is perceived accuracy. It is influenced by users’ skepticism, as heightened skepticism may lead individuals to scrutinize information more critically. Lastly, perceived content diagnosticity is the final mediator in the sequence. As individuals become skilled at identifying valuable and reliable information, they are less likely to view fake news stories as diagnostic or informative. Therefore, media literacy interventions, by enhancing users’ skepticism and decreasing the perceived accuracy of fake news,

contribute significantly to diminishing users’ assessments of fake news content diagnosticity and their propensity to share fake news stories. Thus, to examine the underlying mechanisms that link digital literacy interventions with intentions to engage in online behavior, we hypothesized that:

H5: The difference in intentions to engage in online activities between the control group and (1) a priming critical thinking intervention or (2) an experiential training intervention is serially mediated through users’ skepticism, perceived accuracy of the news item, and perceived content diagnosticity.

Methodology

The objective of the present research was to evaluate whether and how different types of digital literacy interventions reduce belief in false news stories and their proliferation in social media. An additional goal was to explore the underlying mechanisms that link these digital literacy interventions with the perceived accuracy of fake news and online behavioral outcomes. This study was approved by the institutional review board committee of the authors’ academic institution. The research conformed to accepted ethical guidelines. Informed consent was obtained from all respondents, and their anonymity was safeguarded.

The research focused primarily on online content presented in the form of news tweets. An experimental design was used to examine the hypotheses. The experiment was conducted during national electoral campaigns one week before the 2022 Israeli elections (October 2022), a period of high political interest. A news prototype was designed to serve as stimulus materials. The research included a fake news tweet on the housing crisis (see Appendix 1 for details). The fake news stimulus featured Benjamin Netanyahu, then former prime minister of Israel, announcing that Israelis must realize there is no solution to the housing crisis. The news item was a modification of actual fake news items posted on social media. The fake news message employed in the experiment underwent a meticulous selection process involving expert judgment and multiple rounds of pretesting. Given the political nature of the issue, political scientists with a profound understanding of the subject matter and the capacity to evaluate accuracy, tone, and impact within this domain were engaged as experts. Their valuable input during the selection process ensured that the chosen message effectively represented fake news content. In addition, the fake news message was validated in a pretest regarding the realism of content, appropriateness of length, and readability.

We conducted an online panel experiment that included embedded media literacy interventions. Participants were randomly assigned to different types of digital literacy

interventions. Next, they viewed news tweets and responded to questionnaires remotely on their mobile phones. After participating in their randomly assigned literacy interventions and reviewing their (fictional) news tweets, the participants completed the attention check and reported their perceptions of post accuracy and intentions to engage in online activities related to the news content. In addition, they reported their perceptions of skepticism and content diagnosticity. Finally, we collected demographic information from participants regarding their gender, age, education level, and political alignment. At the end of the experiments, we informed participants that the tweet they had seen was a fake news story.

Materials and procedures

Through an online experiment, we tested two conditions and a control. Each participant was randomly assigned to one of three conditions: (1) the non-intervention control condition, in which participants were not exposed to any interventions; (2) the boost-based intervention designed to prime critical thinking; and (3) the experiential training intervention condition designed to immerse social media users in an experience. Next, we presented participants with news items formatted as a Tweeter tweet: a picture (i.e., a graph illustrating the commencement and completion of construction by year) accompanied by a headline and a few sentences.

Participants allocated to the non-intervention control condition were expected to interact with a Tweeter tweet without receiving any explicit guidance or prompts concerning accuracy-related considerations. Initially, the control condition was designed with the intention of being a neutral baseline, free from any deliberate intervention. Its primary objective was to act as a standard against which the effects of other interventions could be measured. However, as participants in the control condition read the tweet, they were subsequently encouraged to assess the accuracy of the selected tweet. This apparently innocuous query functioned as a catalyst, compelling participants to delve into the realm of accuracy-related considerations regarding the presented information. The subtle but significant transformation of a non-intervention control condition into a prompting accuracy intervention is a noteworthy aspect of this study.

In the priming critical thinking intervention condition, we considered the effects of exposure to Facebook's "Tips to Spot False News." These tips are almost certainly the most widely disseminated digital media literacy intervention conducted to date (Guess et al., 2020). The tips, adapted verbatim from Facebook's campaign, consist of 10 strategies readers can use to identify false stories that appear in their news feeds (see Appendix 2 for the full intervention). The tips provide simple rules that can help individuals evaluate the credibility of sources and identify indicators of problematic content. For example, one sample tip recommends that

respondents "[w]atch for unusual formatting," warning that "many false news sites have misspellings or awkward layouts." Significantly, this method's success does not depend on readers taking time-consuming actions such as conducting research or carefully considering every piece of news they encounter. Instead, it aims to provide clear decision-making guidelines that assist people in differentiating real news from fake.

Finally, in the experiential training intervention condition, we employed a short exercise called Fake or Real, adapted from the Israeli Institute for National Security Studies (INSS). Participants were instructed to participate in an exercise for testing the identification of fake photos of people. Ten images were displayed after participants clicked on a link. Some were real, and some fake (created by a machine). The participants were asked to select real or fake for each of the ten images (see Appendix 3 for details). The absence of tips or guidance in the exercise is a deliberate choice that underscores the essence of experiential learning. Experiential learning, rooted in Kolb's theory (Kolb, 2014), posits that knowledge is constructed through direct sensory experiences and contextual actions. By withholding explicit tips, the "Fake or Real" exercise compels participants to rely solely on their existing knowledge, intuition, and critical thinking abilities. This approach encourages active engagement and forces participants to directly confront the challenge, mirroring real-world scenarios where misinformation often lacks clear warning signs.

Measurements

Existing scales were used when available, and, where necessary, slight changes in wording were made to adapt the questions to the study context. The questions appeared in Hebrew. The backward translation method was used. The measurement items are presented in Appendix 4 Table 6.

Perceived accuracy Was measured using a 6-point Likert scale: 1 = *Definitely false*, 2 = *Probably false*, 3 = *Possibly false*, 4 = *Possibly true*, 5 = *Probably true*, 6 = *Definitely true*. This measure was adapted from Martel et al. (2020).

User intention to engage in online activities Was measured using a modified version of Nguyen and Sharkasi's (2021) Intention to Engage on Facebook measure. Participants were asked to indicate on a 7-point Likert-type scale (1 = *very unlikely* and 7 = *very likely*) how likely they were to (1) click the "share" button, (2) click the "link" button to further reading, and (3) comment on their assigned tweet. The three items were aggregated to a mean index. Cronbach's alpha, composite reliability (CR) and average variance extracted (AVE) coefficients for the current study are presented in Table 2.

Users' skepticism Users' pre-dispositional skepticism reflects individuals' innate inclination to approach online information with a degree of doubt and critical evaluation. We evaluated participants' levels of suspicion and concerns regarding posters' motivations and identities. Constructs of fake identity and ulterior motivation were measured using X. J. Zhang et al.'s (2016) electronic word-of-mouth skepticism scale. We modified the items for the context of electronic news messages. The current study used the "fake identity" and "ulterior motivation" scales to evaluate participants' pre-dispositional suspicion toward online news communications. Cronbach's alpha, CR and AVE coefficients for the current study are presented in Table 2.

Fake identity was measured using a modified version of the three-item fake identity scale developed by X. J. Zhang et al. (2016). Sample items include "People write online news posts pretending they are someone else." and "Different online news posts are often posted by the same person under different names." All the items were measured on 7-point Likert scales ranging from "strongly disagree" to "strongly agree."

Ulterior motivation was measured using a modified version of the three-item ulterior motivation scale developed by X. J. Zhang et al. (2016). Sample items include "Most online news posts intend to mislead." and "People writing online news posts are always up to something." All the items were measured on 7-point Likert scales ranging from "strongly disagree" to "strongly agree."

Perceived content diagnosticity Was measured using a modified version of the online customer review diagnosticity scale developed by Jiang and Benbasat (2004, 2007) and Filieri et al., (2018). The scale consisted of four items. Sample items include "The tweet provided valuable information on the candidate's view of the housing crisis." and "The information provided in the tweet helped me to evaluate the candidate's view of the housing crisis." All the items were measured on 7-point Likert scales ranging from "strongly disagree" to "strongly agree." Cronbach's alpha, CR and AVE coefficients for the current study are presented in Table 2.

Participants

The data presented in this research was collected from a national online panel of participants recruited by iPanel, a national internet sample provider. Panel members were randomly sampled from quota groups that matched national census characteristics. Specifically, iPanel's matching and weighting algorithm selected respondents to approximate the demographic and political attributes of the Israeli population. Initially, 6570 people received invitations to participate in this study. Of these, 3701 participants responded, and

1599 participants dropped out before the study ended. We removed 1493 participants from the sample because they failed the attention checks. We were left with a final sample of 609 participants aged 18–70 ($M=41.67$ $SD=15.18$), of which 306 (50.2%) were female and 303 (49.8%) male. Additional background variables are presented in Appendix 5 Table 7. Comparing the three conditions in terms of background variables yielded non-significant results for all variables including political alignment (see Appendix 5 Table 7).

Attention check

In line with Guess et al. (2020), we acknowledge that while we could ask the respondents in the "priming critical thinking" intervention to read the digital literacy fake news tips, we could not force them to read them carefully. Thus, our indicator for receipt of treatment is the ability to correctly answer a series of two follow-up questions about the content of the news tips. In addition, following the recommendations of Berinsky et al. (2014), we added a multiple-choice question regarding the topic of the news tweets to check whether the participants were reading the text. These questions were asked just prior to the news-accuracy task. Participants who did not answer the attentiveness screening questions correctly were omitted from the study.

After participating in their randomly assigned literacy interventions and reviewing their (fictional) news tweet, the participants completed the attention check and reported their perceptions of message accuracy and intentions to engage in online activities related to the news content. In addition, they reported their perceptions of skepticism and content diagnosticity.

Data analysis

In the pilot study, a mixed-design analysis of variance (ANOVA) was employed to assess the effect of condition (control, priming critical thinking, and experiential training) and post type (fake vs. real news) on accuracy. In the main study, we focused on the effect of condition on accuracy and intentions, using exclusively a fake news item. Prior to delving into these analyses, we ensured the reliability and convergent validity of the constructs by calculating Cronbach's alpha, CR, and AVE coefficients. Cronbach's alpha and CR values are deemed satisfactory when they exceed 0.7, indicating good reliability (Hair & Alamer, 2022). Similarly, an AVE value of 0.5 or higher is considered acceptable (Fornell & Larcker, 1981; Hair & Alamer, 2022). The Heterotrait-Monotrait (*HTMT*) ratio was used to assess discriminant validity. The acceptable level of discriminant validity is suggested to be less than 0.90 (Henseler et al., 2015).

In addition, differences in background and study variables by study group were analyzed. One-way ANOVA or Welch's test followed by Tukey's HSD or Games-Howell post-hoc tests were used for continuous variables. χ^2 test for independence was examined for categorical variables. Next, associations between continuous variables were calculated via zero-order correlations. Associations between dichotomous and continuous variables were estimated with point-biserial correlations. Finally, mediation analyses were conducted via one path analysis model. Group types were entered into the model as two dummy variables (priming and experiential) with the control group as a reference. Model fit was evaluated using five goodness of fit indices (Hu & Bentler, 1999): χ^2 statistic, which is considered acceptable when the value is not significant; χ^2/df index, which is considered acceptable when the value is less than five and between one and three an excellent fit; the comparative fit index (CFI), with adequate values above 0.90 and excellent fit of above 0.95; the root mean square error of approximation (RMSEA), with values less than 0.08 as an adequate fit and less than 0.06 as an excellent fit; and standardized root mean squared residual (SRMR), with values less than 0.10 as an adequate fit and less than 0.08 as an excellent fit. In addition, to reach the most parsimonious model, non-significant paths were omitted (Bentler & Mooijaart, 1989). Comparisons between the models in the fit indices were examined using the χ^2 difference test. When the difference between the initial model and the parsimonious model was insignificant, we concluded that the fit indices of the parsimonious model were better than those of the initial model (Kline, 2016). Confidence intervals (CI) were estimated for the indirect effects based on 5000 bootstrap samples of the data. CIs that excluded zero indicated a significant indirect effect (Shrout & Bolger, 2002).

The statistical tests were performed using IBM SPSS Statistics version 28. Path analysis was conducted using IBM SPSS Amos version 24 using the "specific indirect effects" estimand (Gaskin et al., 2020) to analyze the specific indirect effects. The significance level for testing the research hypotheses was 5%.

Results

A pilot study - false news vs. real news content

We conducted a small-scale pilot study before the main research to check the study's feasibility. We aimed to address the concerns that any positive results in participants' ability to identify fake news resulting from exposure to digital literacy interventions would be counterbalanced by encouraging people to distrust real news. Previous research indicated that these positive effects are not merely an artifact of increased skepticism toward all information (Guess et al., 2020). In fact, exposure to the

intervention widened the gap in perceived accuracy between mainstream and false news headlines overall (Guess et al., 2020). Hence, prior to embarking on a comprehensive research project exclusively focused on evaluating the interventions with a fake news tweet, we conducted a preliminary study using real news content as well. We evaluated whether the media literacy interventions improved respondents' ability to distinguish between false news stories and mainstream news content.

The pilot study contained two Tweeter tweets about the housing crisis, one based on fake news and one on real news content (see Appendix 6). For the fake news stimulus, we utilized the same content as in our main study. The real news story featured Benjamin Netanyahu presenting his solution for the housing crisis. In each tweet shown to participants, the date, number of likes, and number of retweets were identical. Participants in each of the three conditions (i.e., control, priming critical thinking, and experiential training) were randomly assigned to view one of the two tweets. After reading the tweets and completing the attention check questions, participants reported their perceptions of their accuracy. The sample for the pilot study consisted of 151 participants (mean age = 41.48 years; age range = 18–70; 51.0% men; 49.0% women; 58.3% college graduates; 45% identified as politically center or left-wing and 46.4% as right-wing [8.6% refused to answer the question on political leanings]).

A mixed-design ANOVA detected a significant main effect of post type on accuracy [$F(1, 148) = 12.27, p < 0.001, \eta^2_p = 0.08$] such that the mean score for real tweets was higher than for fake tweets (see Table 1 and Fig. 1). These results indicate that exposure to media literacy interventions did not encourage people to distrust real news. Fake news stories had lower accuracy rates than real news stories above and beyond all conditions. The main effect of condition was non-significant [$F(2, 148) = 1.06, p = 0.351, \eta^2_p = 0.01$], as was the interaction between post type and condition [$F(2, 148) = 0.50, p = 0.605, \eta^2_p = 0.01$].

The main study - preliminary analysis

Table 2 presents reliability and validity coefficients. Overall, the values fell within the acceptable range of the recommended threshold values.

Table 3 presents associations between background and study variables. As it shows, all intercorrelations between study variables were significant. Skepticism was negatively associated with all the other variables. Accuracy was positively associated with diagnosticity and intentions. Last, diagnosticity was positively associated with intentions. Regarding the background variables, gender was associated with skepticism and accuracy, indicating that women scored lower on skepticism and higher on accuracy than men. Education was associated with skepticism, accuracy, and diagnosticity. Participants with academic educations scored higher on skepticism and lower on accuracy

Table 1 Means and standard deviations of accuracy by post type and condition

Post type	Condition						Effect	F	η^2_p	
	Control		Priming critical thinking		Experiential intervention					Total
	(n = 51)	(n = 50)	(n = 50)	(n = 50)	(n = 151)					
	M	SE	M	SE	M	SE				
Real	3.14	0.20	3.04	0.20	3.00	0.20	T	12.27***	0.08	
Fake	2.69	0.19	2.22	0.20	2.52	0.20	C	1.06	0.01	
Total	2.91	0.14	2.63	0.14	2.76	0.14	T×C	0.50	0.01	

T post type, C condition

*** p < 0.001

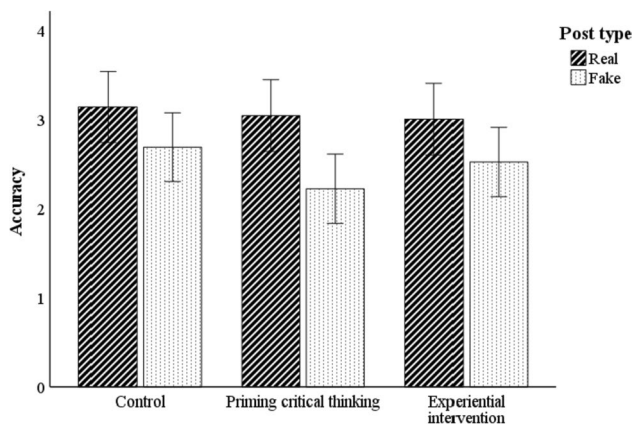


Fig. 1 Means and 95% confidence intervals of accuracy by post type and condition. Note. N = 151

and diagnosticity than those with non-academic educations. Last, income was positively associated with skepticism. Appendix 7 Table 8 presents means and standard deviations of study variables by dichotomous background variables.

It’s worth noting that the associations between political alignment and all study variables were not statistically significant. Additionally, as previously stated, there were no significant differences in political alignment across the three conditions. Consequently, political alignment was not included as a control variable in all analyses.

The main study - hypotheses testing

The hypotheses were tested twice: with the control variables (gender and education) and without them. Since the results were not substantially different, they were reported without controlling them. Table 4 presents comparisons between the groups by study variables. All analyses yielded significant results. Post-hoc tests indicated that the priming critical thinking and experiential intervention groups scored lower on accuracy than the control group, indicating support for H1a and H2a. Regarding intentions to engage in online activities, the priming critical thinking group scored lower than the other two groups, which were not significantly different, thus indicating support for H1b but not for H2b. In addition, the priming critical thinking and experiential intervention groups scored higher on skepticism than the control group, indicating support for H3a.

Path analysis included the hypothesized direct and indirect effects. Additional paths were added to the model as new correlations, not hypothesized, were found in Table 3. The model exhibited excellent fit with the data [$\chi^2(2) = 0.23, p = .892, \chi^2/df = 0.12, CFI = 1.00, RMSEA < 0.001, SRMR = 0.01$]. An examination of the path model showed several non-significant paths (see dashed lines in Fig. 2). Therefore, a more parsimonious model without these insignificant paths was examined. A comparison between the models showed that the χ^2 did not change significantly ($\Delta\chi^2 = 4.79, \Delta df = 5, p = 0.686$), thus making the parsimonious model preferable. The final

Table 2 Reliability and validity coefficients of the constructs

Variable	α	CR	AVE	HTMT	
				1	2
1. Skepticism	.89	.89	.58	—	
2. Diagnosticity	.91	.92	.75	0.31	—
3. Intentions	.71	.74	.49	0.36	0.48

Cronbach’s alpha, CR, AVE, and HTMT values are presented. CR Composite reliability, AVE Average variance Extracted, HTMT Heterotrait-Monotrait

Table 3 Means, standard deviations, and zero-order correlations between background and study variables

Variable	1	2	3	4
1. Skepticism	—			
2. Accuracy	-0.62***	—		
3. Diagnosticity	-0.28***	0.43***	—	
4. Intentions	-0.29***	0.31***	0.39***	—
5. Age	0.04	-0.02	-0.07	0.01
6. Gender ^a	-0.09*	0.13**	0.003	-0.06
7. Education ^b	0.09*	-0.13**	-0.11**	-0.07
8. Income	0.09*	-0.05	0.001	-0.05
9. Political alignment ^c	0.07	-0.01	0.01	-0.05
<i>M</i>	4.74	2.74	2.79	2.36
<i>SD</i>	1.40	1.30	1.53	1.39

N = 609. Zero-order coefficients are presented for continuous variables. For correlations between dichotomous and continuous variables, point-biserial coefficients are presented. ^a0 = Male, 1 = Female; ^b0 = Non-academic, 1 = Academic; ^c0 = Right, 1 = Left

* *p* < 0.05. ** *p* < 0.01. *** *p* < 0.001

model, presented in Fig. 2, displayed an excellent fit with the data [$\chi^2(7) = 5.02, p = .658, \chi^2/df = 0.72, CFI = 1.00, RMSEA < 0.001, SRMR = 0.02$].

The paths from priming critical thinking and experiential interventions to skepticism and intentions were significant, indicating support for *H1b*, *H2b* and *H3a*. The path from skepticism to accuracy was negative and significant, indicating support for *H3b*. In addition, the path from skepticism to intentions was negative and significant. The path from accuracy to diagnosticity was positive and significant, as was the path from diagnosticity to intentions, indicating support for *H4a* and *H4b*, respectively.

Next, the indirect effects were examined (see Table 5). The relative indirect effects of priming critical thinking and experiential intervention groups on accuracy through skepticism were significant, indicating support for *H3c*. Specifically, both the priming critical thinking and experiential intervention groups scored higher than the control group on skepticism, and

higher skepticism scores predicted lower accuracy scores. The indirect effect of accuracy on intentions through diagnosticity was also significant, indicating support for *H4c*. The results revealed that lower accuracy scores predicted lower diagnosticity scores, which subsequently predicted lower intentions to engage in online activities scores. Lastly, the relative indirect effects of priming critical thinking and experiential intervention groups on intentions through skepticism, accuracy, and diagnosticity were significant, indicating support for *H5*. As mentioned, priming critical thinking and experiential intervention groups scored higher than the control group on skepticism. Higher skepticism scores predicted lower accuracy scores, which in turn predicted lower diagnosticity scores, which subsequently predicted lower intention scores. In addition, non-hypothesized, relative indirect effects of priming critical thinking and experiential intervention groups on intentions through skepticism were found.

Discussion

The current study sought to assess the impact of digital literacy interventions on participants' ability to identify unreliable information in news articles. Specifically, the study aimed to evaluate the effectiveness of experience-based (i.e., training in identifying whether photos are fake or real) and boost-based priming critical thinking (i.e., reading guidelines for evaluating news online) digital literacy interventions.

Theoretical conclusions

This research makes several theoretical contributions to the study of digital literacy interventions. First, it contributes to the growing body of research (see Kozyreva et al., 2022) aimed at reducing people's vulnerability to fake news by introducing a new experiential training digital literacy intervention strategy. Our results show that a brief intervention that could be inexpensively disseminated at scale can effectively reduce users'

Table 4 Means and standard deviations of the study variables by group

Variable	Group						<i>F</i>	η^2 / ω^2
	Control		Priming critical thinking		Experiential intervention			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Skepticism	4.43 _a	1.45	4.91 _b	1.26	4.89 _b	1.43	8.02***	0.03
Accuracy	3.00 _a	1.29	2.56 _b	1.20	2.65 _b	1.37	6.75**	0.04
Intentions	2.60 _a	1.47	2.05 _b	1.17	2.41 _a	1.47	9.28***	0.05

ANOVA *F* and eta-squared are presented for skepticism. For accuracy and intentions, Welch's *F* and omega-squared are presented. Means with different subscript letters differ significantly at the 0.05 level

** *p* < 0.01. *** *p* < 0.001

perceived accuracy of fake news stories. Second, our findings confirm that the priming critical thinking and experiential training media literacy interventions significantly impacted the perceived accuracy of fake news and users' subsequent online activities. Finally, this research deepens our understanding of the underlying mechanisms that influence customers' accuracy judgments and subsequent online behavioral intentions. Specifically, following Guess et al.'s (2020) recommendation to study the underlying mechanisms through which different digital literacy interventions operate, our conceptual contribution involves the formulation of a model that links literacy interventions to perceived accuracy and online behavioral outcomes through the elicitation of users' skepticism and perceived content diagnosticity.

It is important to note that because it asks the participants to indicate a post's accuracy, the control condition is, in fact, a prompting accuracy intervention condition. In other words, this group initially functioned as a non-intervention control condition since there was no intervention prior to participants' viewing of a Tweeter tweet. However, after reading the tweet, participants were asked if the selected tweet was accurate. By asking this question, we prompted participants to consider an accuracy motive. As a result, this non-intervention condition turned into a prompting accuracy intervention, and this fact may explain why the accuracy and the intentions to engage in online activities scores are relatively low even for the control condition.

Our findings demonstrate that the underlying mechanism that links digital literacy interventions with perceived accuracy and intentions to engage in online activities varies depending on the type of digital literacy intervention. Both the boost-based priming critical thinking intervention and the experiential training intervention scored higher than the control group (i.e., a nudge-based prompting accuracy intervention) on skepticism, and higher skepticism scores predicted lower accuracy scores. In addition, in support of our hypotheses, the results revealed that lower accuracy scores predicted lower diagnosticity scores, which subsequently predicted lower intentions to engage in online activities scores. Finally, serial mediation analyses showed that users' skepticism, perceived accuracy, and perceived content diagnosticity mediate the association between digital literacy interventions and intentions to engage in online activities. Specifically, our findings demonstrate that, in line with our hypotheses, the relative indirect effects of priming critical thinking and experiential intervention on intentions through skepticism, accuracy, and diagnosticity were significant. These findings may contribute to a better understanding of the mechanisms through which various digital literacy interventions operate. We can conclude that the experiential training intervention may genuinely target competencies rather than simply prompting accuracy motivation. Indeed, although, unlike other boost-based interventions, the experiential training intervention is quick and does not require users' active engagement with reading tips, this intervention improves respondents'

ability to identify fake news content and lowers their intentions to engage in online activities related to fake news.

Managerial implications

Fundamentally, digital literacy interventions designed to equip people with media literacy skills have the potential to build competencies and thus reduce the amount of fake news information that circulates online. The present research tested an innovative strategy to improve social media users' ability to identify fake news content through an experiential training approach that allowed them to directly experienced the task of distinguishing between true and fake content. By challenging participants to identify fake images without explicit tips, this exercise not only tests their current abilities but also encourages the development of critical thinking skills, fostering a sense of curiosity and skepticism. Participants are actively engaged in the exercise, which enhances their cognitive involvement and fosters a spirit of inquiry. Instead of passively receiving information, they become co-creators of knowledge. Moreover, since experiential training interventions involve a hands-on approach, participants remained actively engaged throughout the training and were thus likely to internalize and recall it after the training session ended. Our quick exercise intervention could easily translate into social media platform interventions to increase users' focus on critical thinking and accuracy. Media sites could, for instance, regularly offer users the suggested brief exercise to subtly and effectively remind them about accuracy. Since this intervention is brief and game-like, media users will probably be amenable to participating in it.

Limitations and future research

In light of the growing interest in the impact of digital literacy interventions (see Pennycook et al., 2021), our findings suggest fruitful avenues for further research on consumers' responses to fake news. Several future research directions stem from the limitations inherent in this research. First, a notable limitation in our main study lies in the restricted focus on assessing the effectiveness of digital literacy interventions using solely a fake news item. While previous studies have explored the impact of interventions encompassing both true and false headlines, our examination of individuals' discernment between true and false news items was confined to the pilot study. To address this deficiency, future research should endeavor to conduct comprehensive comparative studies that encompass both true and false news items. Such an approach would provide researchers with a more comprehensive understanding of how distinct interventions influence individuals' capacity to distinguish between reliable and unreliable news content.

Second, the generalizability of our findings may be constrained by the experiment's concentration on a specific topic (i.e., the housing crisis). To enhance the applicability of our

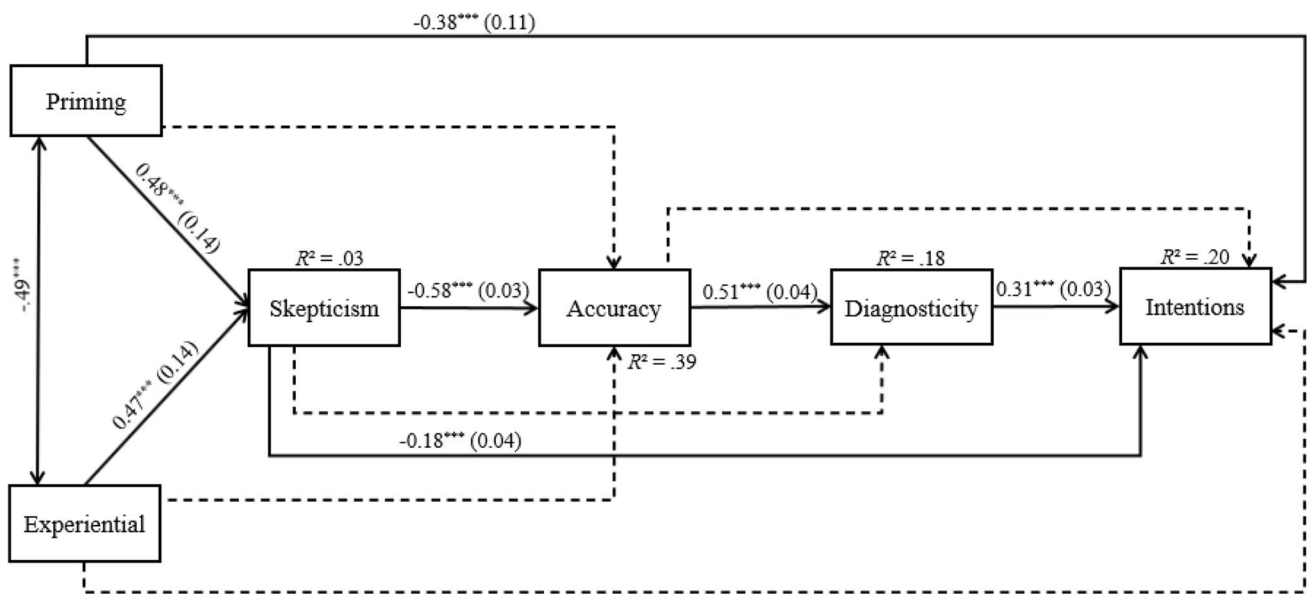


Fig. 2 Path analysis model. Note. $N=609$. Values are unstandardized regression coefficients. In parentheses: standard errors. Solid lines indicate significant paths, and dashed lines indicate non-significant

trimmed paths. Priming = Priming critical thinking vs. control; Experiential = Experiential intervention vs. control. *** $p < 0.001$

Table 5 Bootstrap results for the indirect effects

Indirect effect	B	SE	95% Bootstrap CI	
			Lower	Upper
Priming → Skepticism → Accuracy	-0.28	0.08	-0.43	-0.13
Experiential → Skepticism → Accuracy	-0.27	0.09	-0.44	-0.10
Accuracy → Diagnosticity → Intentions	0.15	0.03	0.10	0.21
Priming → Skepticism → Accuracy → Diagnosticity → Intentions	-0.04	0.01	-0.07	-0.02
Experiential → Skepticism → Accuracy → Diagnosticity → Intentions	-0.04	0.02	-0.08	-0.02
Priming → Skepticism → Intentions	-0.09	0.03	-0.16	-0.03
Experiential → Skepticism → Intentions	-0.09	0.04	-0.16	-0.03

$N=609$, CI Confidence Interval. Priming priming critical thinking vs. control, Experiential experiential intervention vs. Control

results, future investigations should contemplate extending our findings to other online contexts featuring fake news. Exploring the effectiveness of interventions in various online contexts and across different demographic groups can provide valuable insights into tailoring interventions for specific audiences and situations. Additionally, it is crucial to consider motivational factors, such as altruism and status seeking. Which could potentially impact individuals’ sharing behavior (Omar et al., 2023). In our study, all participants engaged in a hypothetical scenario involving fake news related to the housing crisis, likely yielding relatively homogeneous motivational states for the task. Future studies could explore potential moderating effects of motivational factors by varying their levels while individuals process fake news content. This approach would offer a more nuanced understanding of how motivation influences the effectiveness of digital literacy interventions.

Third, prior research has shown that the effects of experimental treatments decline quickly over time (Guess et al., 2020). Since our participants were exposed to the interventions immediately before receiving a fake news post, we cannot assume the interventions would have continued to have an impact if more time had passed between the intervention and the stimulus. Therefore, future research should examine the durability of our interventions’ effects by leveraging a two-wave panel design.

To conclude, the current research explored how social media users reacted to different digital literacy interventions and demonstrated how exposure to an innovative, scalable literacy intervention influenced respondents’ ability to identify fake news content. Moreover, our findings emphasize that developing effective interventions against fake news depends on understanding the underlying mechanisms through which different digital literacy interventions operate.

Appendix 1 The fake news tweet used in the study

Fake news tweet: If there is no bread, let them eat cake...

At the contractors' conference that took place yesterday in a hall in Holon, Benjamin Netanyahu was asked about the rise in housing prices and said: "There is no choice. People must understand that there is no possibility of buying an apartment in the next decade. It's time to say this clearly. People must give up on the dream of buying an apartment and realize that living in a rental is the only solution."



Appendix 2 The priming critical thinking intervention treatments: Facebook "Tips to Spot False News"

1. **Be skeptical of headlines.** False news stories often have catchy headlines in all caps with exclamation points. If shocking claims in the headline sound unbelievable, they probably are.

2. **Look closely at the URL.** A phony or look-alike URL may be a warning sign of false news. Many false news sites mimic authentic news sources by making small changes to the URL. You can go to the site to compare the URL to established sources.
3. **Investigate the source.** Ensure that the story is written by a source that you trust with a reputation for accuracy. If the story comes from an unfamiliar organization, check their "About" section to learn more.
4. **Watch for unusual formatting.** Many false news sites have misspellings or awkward layouts. Read carefully if you see these signs.
5. **Consider the photos.** False news stories often contain manipulated images or videos. Sometimes the photo may be authentic, but taken out of context. You can search for the photo or image to verify where it came from.
6. **Inspect the dates.** False news stories may contain timelines that make no sense, or event dates that have been altered.
7. **Check the evidence.** Check the author's sources to confirm that they are accurate. Lack of evidence or reliance on unnamed experts may indicate a false news story.
8. **Look at other reports.** If no other news source is reporting the same story, it may indicate that the story is false. If the story is reported by multiple sources you trust, it's more likely to be true.
9. **Is the story a joke?** Sometimes false news stories can be hard to distinguish from humor or satire. Check whether the source is known for parody, and whether the story's details and tone suggest it may be just for fun.
10. **Some stories are intentionally false.** Think critically about the stories you read, and only share news that you know to be credible.


[These tips are taken verbatim from the original tips published by Facebook (see <https://www.facebook.com/help/188118808357379>).]

Appendix 3 The experiential training intervention: the fake or real exercise adopted from the Israeli Institute for National Security Studies (INSS)

Instructions: In the link in front of you, there is an exercise for testing the identification of fake photos. After you click on the link, ten images will be displayed in front of you. Some are real, and some are fake (created by a machine). Please select real (green) or fake (red) as quickly as possible.

Welcome to Fake or Real!


We will present you 10 photos. Some of them are real and others are fake (created by a machine). Please choose real or fake as fast as you can.



Fake Real

Start

○ ○ ○ ○ ○ ○ ○ ○ ○ ○



Fake Real

🕒 00:09 0 Points

Appendix 4

Table 6 Measurements

Constructs	Items	Source
Skepticism	Fake identity	Adopted from eWOM Skepticism scale (X. J. Zhang et al., 2016)
	1. People writing online news posts are not necessarily the real news source.	
	2. People write online news posts pretending they are someone else.	
	3. Different online news posts are often posted by the same person under different names.	
	Ulterior motivation	
	1. Online news posters care more about influencing you than providing reliable information.	
Perceived post diagnosticity	1. The tweet provided valuable information on the candidate's view of the housing crisis.	Adopted from review diagnosticity (Jiang & Benbasat, 2004; Filieri et al., 2018)
	2. The information provided in the tweet helped me evaluate the candidate's view of the housing crisis.	
	3. The information provided in the tweet helped me gain familiarity with the candidate's view of the housing crisis.	
	4. The information provided in the tweet helped me understand the candidate's positions on the housing crisis.	
Intentions to engage in online activities	1. Clicking a link for further reading	
	2. Commenting on the post	
	3. Sharing with people	

Appendix 5

Table 7 Background characteristics

Variable	Total sample (<i>N</i> = 609)		Conditions			<i>F</i>	η^2			
			Control (<i>n</i> = 206)		Priming critical thinking (<i>n</i> = 202)			Experiential intervention (<i>n</i> = 201)		
A. Continuous	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	0.09	< 0.001
Age	41.67	15.18	41.31	15.40	41.91	15.10	41.79	15.10		
Income	3.04	1.39	3.02	1.46	3.01	1.34	3.10	1.38	0.24	0.001
B. Categorical	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	χ^2	0.09
Gender										
Male	303	49.8%	101	49.0%	102	50.5%	100	49.8%		
Female	306	50.2%	105	51.0%	100	49.5%	101	50.2%		
Education									3.95	
Non-academic	297	48.8%	107	51.9%	87	43.1%	103	51.2%		
Academic	312	51.2%	99	48.1%	115	56.9%	98	48.8%		
Political alignment (Elections 2021)									0.70	
Right	274	49.1%	97	51.1%	87	46.8%	90	49.5%		
Left	284	50.9%	93	48.9%	99	53.2%	92	50.5%		

Data were missing for 51 cases for political alignment. Sixty-six participants refused to answer the question regarding income

Appendix 6 The fake news and real news tweets used in the pilot study

Real news tweet: A substantial and simple solution to the housing crisis...

At the contractors’ conference that took place yesterday in a hall in Holon, Benjamin Netanyahu was asked about the rise in housing prices and said: “There are 300,000 housing units designated for accommodating 1.2 million residents. However, delays have arisen due to negotiations involving government authorities and the budget

department. I pledge to lead a housing cabinet dedicated to swiftly releasing these units, offering a timely solution to the housing challenge.”

Fake news tweet: If there is no bread, let them eat cake...

At the contractors’ conference that took place yesterday in a hall in Holon, Benjamin Netanyahu was asked about the rise in housing prices and said: “There is no choice. People must understand that there is no possibility of buying an apartment in the next decade. It’s time to say this clearly. People must give up on the dream of buying an apartment and realize that living in a rental is the only solution.”

Appendix 7

Table 8 Means and standard deviations of study variables by dichotomous background variables

Variable	Gender		Education		Political alignment	
	Male (n = 303)	Female (n = 306)	Non-academic (n = 297)	Academic (n = 312)	Right (n = 274)	Left (n = 284)
Skepticism	4.87	4.61	4.61	4.86	4.61	4.82
	1.38	1.40	1.35	1.43	1.41	1.40
Accuracy	2.57	2.91	2.91	2.58	2.76	2.73
	1.35	1.23	1.26	1.32	1.37	1.26
Intentions	2.45	2.27	2.45	2.26	2.44	2.30
	1.41	1.37	1.50	1.28	1.38	1.40

Each cell represents means (above) and standard deviations (below)

Data availability All data and research materials are available at [<https://osf.io/ukbe4>; identifier: <https://doi.org/10.17605/OSF.IO/UKBE4>].

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Ahadzadeh, A. S., Ong, F. S., & Wu, S. L. (2023). Social media skepticism and belief in conspiracy theories about COVID-19: The moderating role of the dark triad. *Current Psychology*, *42*, 8874–8886. <https://doi.org/10.1007/s12144-021-02198-1>

Altay, S., De Araujo, E., & Mercier, H. (2022). “If this account is true, it is most enormously wonderful”: Interestingness-if-true and the sharing of true and false news. *Digital Journalism*, *10*(3), 373–394. <https://doi.org/10.1080/21670811.2021.1941163>

Amazeen, M. A., Thorson, E., Muddiman, A., & Graves, L. (2018). Correcting political and consumer misperceptions: The effectiveness and effects of rating scale versus contextual correction formats. *Journalism & Mass Communication Quarterly*, *95*(1), 28–48. <https://doi.org/10.1177/1077699016678186>

Amsalem, E., & Zoizner, A. (2023). Do people learn about politics on social media? A meta-analysis of 76 studies. *Journal of Communication*, *73*(1), 3–13. <https://doi.org/10.1093/joc/jqac034>

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. <https://doi.org/10.1037/xge0000729>

Basol, M., Roozenbeek, J., & Van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, *3*(1), 1–9. <https://doi.org/10.5334/joc.91>

Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin*, *106*(2), 315–317. <https://doi.org/10.1037/0033-2909.106.2.315>

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, *58*(3), 739–753. <https://doi.org/10.1111/ajps.12081>

- Bigne, E., Ruiz, C., Cuenca, A., Perez, C., & Garcia, A. (2021). What drives the helpfulness of online reviews? A deep learning study of sentiment analysis, pictorial content and reviewer expertise for mature destinations. *Journal of Destination Marketing & Management*, 20. <https://doi.org/10.1016/j.jdmm.2021.100570>
- Brodsky, J. E., Brooks, P. J., Scimeca, D., Todorova, R., Galati, P., Batson, M., Grosso, R., Matthews, M., Miller, V., & Caulfield, M. (2021). Improving college students' fact-checking strategies through lateral reading instruction in a general education civics course. *Cognitive Research: Principles and Implications*, 6(1). <https://doi.org/10.1186/s41235-021-00291-4>
- Cook, J., Lewandowsky, S., & Ecker, U. K. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLoS One*, 12(5), e0175799. <https://doi.org/10.1371/journal.pone.0175799>
- Duffy, A., & Ling, R. (2020). The gift of news: Phatic news sharing on social media for social cohesion. *Journalism Studies*, 21(1), 72–87. <https://doi.org/10.1080/1461670X.2019.1627900>
- Duncan, M. (2022). What's in a label? Negative credibility labels in partisan news. *Journalism & Mass Communication Quarterly*, 99(2), 390–413. <https://doi.org/10.1177/1077699020961856>
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Edelman (2021). Edelman Trust Barometer. Available: <https://www.edelman.com/trust/2021-trust-barometer>. Accessed 30 Jul 021
- Eshet-Alkalai, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, 13(1), 93–106.
- Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School (HKS) Misinformation Review*, 1(2), 1–8. <https://doi.org/10.37016/mr-2020-009>
- Feuerstein, M. (1999). Media literacy in support of critical thinking. *Journal of Educational Media*, 24(1), 43–54. <https://doi.org/10.1080/1358165990240104>
- Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information & Management*, 55(8), 956–970. <https://doi.org/10.1016/j.im.2018.04.010>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.2307/3151312>
- Gaskin, J., James, M., & Lim, J. (2020). *Specific indirect effects*. AMOS Estimand. Gaskination's StatWiki.
- 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>
- Hair, J., & Alamer, A. (2022). Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research Methods in Applied Linguistics*, 1(3), <https://doi.org/10.1016/j.rmal.2022.100027>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Ireton, C., & Posetti, J. (2018). *Journalism, fake news & misinformation: Handbook for journalism education and training*. Unesco Publishing.
- Jiang, Z., & Benbasat, I. (2004). Virtual product experience: Effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping. *Journal of Management Information Systems*, 21(3), 111–147. <https://doi.org/10.1080/07421222.2004.11045817>
- Jiang, Z., & Benbasat, I. (2007). The effects of presentation formats and task complexity on online consumers' product understanding. *MIS Quarterly*, 475–500. <https://doi.org/10.2307/25148804>
- Jungherr, A., & Schroeder, R. (2021). Disinformation and the structural transformations of the public arena: Addressing the actual challenges to democracy. *Social Media + Society*, 7(1), 1–13. <https://doi.org/10.1177/2056305121988928>
- Kim, B., Xiong, A., Lee, D., & Han, K. (2021). A systematic review on fake news research through the lens of news creation and consumption: Research efforts, challenges, and future directions. *PLoS One*, 16(12), e0260080. <https://doi.org/10.1371/journal.pone.0260080>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT Press.
- Kozyreva, A., Lorenz-Spreen, P., Herzog, S., Ecker, U., Lewandowsky, S., & Hertwig, R. (2022). Toolbox of interventions against online misinformation and manipulation. *PsyArXiv*. <https://doi.org/10.31234/osf.io/x8ejt>
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., ..., Zaragoza, M. S. (2020). *The debunking handbook 2020*. <https://sks.to/db2020>. <https://doi.org/10.17910/b7.1182>
- Li, M., Huang, L., Tan, C. H., & Wei, K. K. (2013). Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce*, 17(4), 101–136. <https://doi.org/10.2753/JEC1086-4415170404>
- Liedke, J. & Matsa, K. E. (2022). Social media and news fact sheet. *Pew Research Center*. [cited 2023 Oct 10]. Available from: <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/#who-consumes-news-on-each-social-media-site>
- Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change*, 58, 101964. <https://doi.org/10.1016/j.gloenvcha.2019.101964>
- Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, 5(1). <https://doi.org/10.1186/s41235-020-00252-3>
- Modirrousta-Galian, A., & Higham, P. A. (2023). Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology: General*, 152(9), 2411–2437. <https://doi.org/10.1037/xge0001395>
- Mudambi, S., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34(1), 185–200. <https://doi.org/10.2307/20721420>
- Nguyen, D. N., & Sharkasi, N. (2021, May). Towards an Understanding of the Intention to Engage on Facebook. In *Advances in Digital Marketing and eCommerce: Second International Conference, 2021* (pp. 62–73). Springer International Publishing.
- Omar, B., Apuke, O. D., & Nor, Z. M. (2023). The intrinsic and extrinsic factors predicting fake news sharing among social media users:

- The moderating role of fake news awareness. *Current Psychology*. <https://doi.org/10.1007/s12144-023-04343-4>
- Panizza, F., Ronzani, P., Martini, C., Mattavelli, S., Morisseau, T., & Motterlini, M. (2022). Lateral reading and monetary incentives to spot disinformation about science. *Scientific Reports*, *12*(1). <https://doi.org/10.1038/s41598-022-09168-y>
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, *66*(11), 4944–4957. <https://doi.org/10.1287/mnsc.2019.3478>
- Pennycook, G., McPhetres, J., Zhang, Y., & Rand, D. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention. *Psychological Science*, *31*, 770–780. <https://doi.org/10.31234/osf.io/uhbk9>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. <https://doi.org/10.1038/s41586-021-03344-2>
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), 1–10. <https://doi.org/10.1057/s41599-019-0279-9>
- Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34). <https://doi.org/10.1126/sciadv.abo6254>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422–445. <https://doi.org/10.1037/1082-989X.7.4.422>
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Wiley-Blackwell.
- Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition*, *133*(3), 572–585. <https://doi.org/10.1016/j.cognition.2014.08.006>
- Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008. <https://doi.org/10.1002/gch2.201600008>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Zhang, X. J., Ko, M., & Carpenter, D. (2016). Development of a scale to measure skepticism toward electronic word-of-mouth. *Computers in Human Behavior*, *56*, 198–208.
- Zhang, L., Wu, L., & Mattila, A. S. (2016). Online reviews: The role of information load and peripheral factors. *Journal of Travel Research*, *55*(3), 299–310. <https://doi.org/10.1177/0047287514559032>
- Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, *53*(5), 1–40. <https://doi.org/10.1145/3395046>
- Zhuang, W., Zeng, Q., Zhang, Y., Liu, C., & Fan, W. (2023). What makes user-generated content more helpful on social media platforms? Insights from creator interactivity perspective. *Information Processing & Management*, *60*(2). <https://doi.org/10.1016/j.ipm.2022.103201>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.