Check for updates

The media literacy dilemma: can ChatGPT facilitate the discernment of online health misinformation?

Wei Peng[1]*, Jingbo Meng[2] and Tsai-Wei Ling[1]

[1]Department of Media and Information, Michigan State University, East Lansing, MI, United States,
[2]School of Communication, The Ohio State University, Columbus, OH, United States

Online health misinformation carries serious social and public health implications. A growing prevalence of sophisticated online health misinformation employs advanced persuasive tactics, making misinformation discernment progressively more challenging. Enhancing media literacy is a key approach to improving the ability to discern misinformation. The objective of the current study was to examine the feasibility of using generative AI to dissect persuasive tactics as a media literacy scaffolding tool to facilitate online health misinformation discernment. In a mixed 3 (media literacy tool: control vs. National Library of Medicine [NLM] checklist vs. ChatGPT tool) × 2 (information type: true information vs. misinformation) × 2 (information evaluation difficulty: hard vs. easy) online experiment, we found that using dissecting persuasive strategies of ChatGPT can be equally effective when compared with the NLM checklist, and that information type was a significant moderator such that the ChatGPT tool was more effective in helping people identify true information than misinformation. However, the ChatGPT tool performed worse than control in terms of helping people discern misinformation. No difference was found in terms of perceived usefulness and future use intention of the ChatGPT tool and the NLM checklist. The results suggest that more interactive or conversational features might enhance usefulness of ChatGPT as a media literacy tool.

Introduction

Online health misinformation is defined as "health-related information disseminated on the Internet that is false, inaccurate, misleading, biased, or incomplete, which is contrary to the consensus of the scientific community based on the best available evidence" (Peng et al., 2023, p. 2133). Such misinformation is different from disinformation in that the creator of the information may not intentionally attempt to make it false or misleading. Online health misinformation carries serious social and public health implications (Romer and Jamieson, 2020; Roozenbeek et al., 2020). Previous studies have endeavored to incorporate techniques for identifying misinformation, including fact-checking labels (Zhang et al., 2021) and warning labels (Pennycook et al., 2020), and media literacy programs with a focus on identifying visual cues (e.g., layouts) or heuristic cues (e.g., sources) (Guess et al., 2020; Vraga et al., 2022). Although these strategies were found effective, a growing prevalence of sophisticated online health misinformation employs advanced persuasive tactics, such as enriching narrative elements, emphasizing uncertainty, evoking emotional responses, and drawing biased conclusions (Peng et al., 2023; Zhou et al., 2023). The next generation of media literacy

programs needs to help the audience fathom the persuasive tactics and critically analyze content and arguments.

Artificial intelligence (AI) and chatbots recently emerged as tools for automatic fact-checking (Guo et al., 2022) or assisting the analysis of arguments and persuasive tactics for misinformation discernment (Altay et al., 2022; Musi et al., 2023). This line of research is still in its nascence and needs more empirical support. Additionally, these tools may suffer from challenges in trust in conversational agents (Rheu et al., 2021) and psychological reactance (Reynolds-Tylus, 2019), rendering them ineffective; that is, people may have little trust in the analysis from AI, even if the analysis is accurate (Choudhury and Shamszare, 2023), and the advice from AI may make people feel irritated due to the threat to their freedom of independent thinking (Pizzi et al., 2021). These conflicting perspectives raise important questions about the feasibility of using ChatGPT as a media literacy tool. Therefore, the objective of the current study was to explore whether ChatGPT can effectively dissect persuasive strategies to support online health misinformation discernment, while also considering the potential limitations of this approach.

Misinformation and persuasive strategies

Infodemic, or the abundance of false or misleading information spreading rapidly through social media and other outlets, intensified during the COVID-19 pandemic and is becoming a global public health issue (Zarocostas, 2020). Various factors contribute to infodemic, including the contemporary media environments with echo chambers and social media filter bubbles (Flaxman et al., 2016); individual factors such as cognitive abilities and biases, political identity, and media literacy level (Nan et al., 2022); and information factors such as the persuasive strategies used to craft the misinformation (Peng et al., 2023). The current study attempts to examine how to improve health misinformation discernment through the angle of understanding persuasive strategies.

Twelve groups of persuasive strategies in online health misinformation were identified in a systematic review of published articles, including content analyses or discourse analyses of information factors that are prevalent in misinformation, and experiments that tested informational factors rendering people vulnerable to misinformation (Peng et al., 2023). These persuasive strategies include fabricating misinformation via vivid storytelling (Peng et al., 2023); using personal experience or anecdotes rather than scientific findings as evidence (Kearney et al., 2019); discrediting government or pharmaceutical companies (Prasad, 2022); making health issues political through the rhetoric of freedom and choice, us vs. them, or religious faith, moral values, or ideology (DeDominicis et al., 2020); highlighting unknown risk and uncertainty (Ghenai and Mejova, 2018); attacking science by exploiting its innate limitation (Peng et al., 2023); inappropriately using scientific evidence to support a false claim (Gallagher and Lawrence, 2020); exaggeration and selectively presentation or omission of information (Salvador Casara et al., 2019); making a conclusion based on biased reasoning (Kou et al., 2017); using fear or anger appeals in persuasion (Buts, 2020); using certain linguistic intensifiers to highlight the points (Ghenai and Mejova, 2018); and establishing the legitimacy of false claims by using certain cues to activate credibility heuristics of people (e.g., medical jargon, seemingly credible source) (Haupt et al., 2021).

Media literacy education to improve misinformation discernment

Combating misinformation consists of two primary lines of research. One is debunking misinformation via fact-check or correction (Chan et al., 2017; Walter et al., 2021). The other line of research is media literacy intervention to improve individuals' capabilities in searching, analyzing, and critically evaluating information (Hobbs, 2010). Media literacy education can help individuals effectively search for information and discover credible sources, interpret and evaluate the information through a critical lens, and be aware of one's biases. A recent meta-analysis demonstrated that media literacy intervention is an effective tool to improve misinformation discernment (Lu et al., 2024). This meta-analysis also revealed that intervention time (ranged from a few minutes to 8 weeks) was not a moderator for the effect size of the outcomes of credibility assessment or attitude related to misinformation, meaning that shorter duration interventions (Guess et al., 2020; Qian et al., 2023) had a similar effect size as long-term interventions (Mingoia et al., 2019; Zhang et al., 2022). Therefore, short and focused interventions with variability to address the fast-paced media environment are advocated (Lu et al., 2024).

Technology-enhanced media literacy tools

What is noticeable is that the meta-analysis demonstrated a moderating effect of the delivery form of the media literacy intervention (Lu et al., 2024). Among the four different forms–course, video, graphic, and game, the *post-hoc* analysis revealed that game-based media literacy interventions (Basol et al., 2021; Roozenbeek and van der Linden, 2019) generally have a larger effect size in terms of assessment of misinformation. This larger effect size can be explained by engaging participants to learn the manipulativeness or persuasive strategies in misinformation via actively playing a role of misinformation creator. For instance, in the *Bad News* game (Roozenbeek and van der Linden, 2019), the participant's goal was to produce news articles using persuasive strategies, such as making a topic look either small and insignificant or large and problematic, or communicating the conspiracy theories to the audience to distrust mainstream narrative. Similarly, in the *Go Viral* game (Basol et al., 2021), the participant's goal was to create emotionally evocative social media posts, or use fake experts to back up the claim, all of which are manipulations through persuasive strategies. This prior evidence of the success of game-based media literacy intervention demonstrated that learning about persuasive strategies can be effective in improving misinformation discernment.

Another type of tools to teach about persuasive strategies may be AI-based tools to prompt the audience to be aware of the persuasive strategies used in the encountered information. AI-based fact-checking or *post-hoc* correction tools have been used extensively in the first line of reach to combat misinformation (Guo et al., 2022). More recently, chatbots emerged as a media literacy tool to assist the analysis of arguments for misinformation discernment (Altay et al., 2022; Musi et al., 2023; Zhao et al., 2023). For instance, the chatbot delivering valid counter-argument was able to move people into a more positive attitude toward GMOs (Altay et al., 2022). The *Fake New Immunity Chatbot* (Musi et al., 2023) interactively taught people to recognize valid arguments and fallacies through reason-checking. Specifically,

the *Fake New Immunity Chatbot* scaffolds the investigation of the connections of the claims and their evidential context by nudging the users into asking critical questions and identifying fallacies such as cherry-picking and false analogy.

These early studies of chatbot-based approaches for combating misinformation have demonstrated feasibility. However, empirical evidence of their effectiveness is lacking. Prior media literacy studies demonstrated that the game-based interventions were particularly successful due to their ability to interactively teach persuasive strategies. The early studies of the chatbot-based approach also alluded that revealing biased reasoning, one of the persuasive strategies commonly used in misinformation, was a promising media literacy approach to combat misinformation, as operationalized as people's attitude toward misinformation or credibility assessment of misinformation. ChatGPT, one of the most widely used large language model (LLM)-based chatbots, has the potential to iteratively reveal persuasive strategies in the information people encounter. Flagging the persuasive strategies employed in the information to the users also serves as an interactive media literacy approach, i.e., teaching users how persuasive strategies are used and increasing their critical thinking. Currently, no empirical study is available to examine the potential of using ChatGPT as a media literacy tool to dissect persuasive strategies (termed as ChatGPT tool thereafter). Therefore, the present study attempts to fill this gap by adding empirical evidence for feasibility. The benchmark to be compared to is a simple media literacy checklist provided by the National Library of Medicine (NLM) as well as a control group. We propose:

> RQ1: How will the ChatGPT tool compare to (a) the NLM media literacy checklist or (b) the control group for users' accuracy in information credibility assessment?

The ChatGPT tool may be effective, but skepticism about AI-powered technology, especially in the field of misinformation detection or correction, is one challenge. The skepticism is rooted in multiple factors, including fears of AI, especially the bias in AI algorithms (De Vito et al., 2017; Zhan et al., 2023), and a lack of understanding about how these systems operate (O'Shaughnessy et al., 2023). Moreover, incidents where AI has amplified misinformation (Zhou et al., 2023) can further erode confidence. Additionally, although flagging persuasive strategies may teach media literacy, people may not like the fact that they are being told by AI what to think and psychological reactance may arise (Pizzi et al., 2021). Therefore, we explore the following research questions to examine whether people accept the ChatGPT tool for assisting information evaluation.

> RQ2: Will participants rate the ChatGPT tool more useful than the NLM checklist for information evaluation?

> RQ3: Will participants be more likely to use the ChatGPT tool than the NLM checklist for future information evaluation?

Method

Study design and procedure

A 3 (media literacy tool: control vs. NLM checklist vs. ChatGPT tool) × 2 (information type: true information vs. misinformation) × 2 (information evaluation difficulty: hard vs. easy) online experiment was employed on *Qualtrics* for randomization. The main independent variable for hypothesis testing was the media literacy tool, a between-subjects variable. The other two independent variables were within subjects. Both true and misinformation were included to control for false positives—the tendency of false skepticism of all information, including true information. Information evaluation difficulty was included to explore whether the media literacy tool works for both simple-to-detect misinformation and well-crafted misinformation.

After giving consent, participants' general information literacy was assessed. Then, they were randomly assigned to one of the three media literacy tool conditions. Except in the control, participants were introduced to how those media literacy tools work and examined their comprehension of these tools, and the correct answers were provided to the participants to reinforce their understanding. Then, all participants viewed four pieces of randomly displayed information (two pieces of true information and two pieces of misinformation) identified from our pilot study. After reading each piece, they were asked about message credibility and issue importance. Participants in the ChatGPT tool and NLM checklist conditions also answered questions about perceived usefulness and future use intention. Participants in the ChatGPT condition were also evaluated on their comprehension of ChatGPT dissecting persuasive strategies, ChatGPT's dissecting of the persuasive strategies, and ChatGPT's dissecting of persuasive strategies.

Participants

A total of 153 participants completed the online experiment. After removing one who failed two of the four attention check tests, 12 who failed the comprehension check of dissecting persuasive strategies of ChatGPT, and one who failed both, 139 were included in the analysis. In total, 52 were in the control condition, 56 in the NLM checklist condition, and 31 in the ChatGPT tool condition. There were fewer participants in the ChatGPT condition due to the removal of participants failing the comprehension and attention check. Among them, 42% ($n = 58$) identified as male, 56% ($n = 78$) as female, 4% ($n = 3$) did not disclose gender; 73% ($n = 101$) identified as White, 15% ($n = 21$) as Black or African American, 7% ($n = 10$) as Asian, 2% ($n = 3$) as mixed race, and 3% ($n = 4$) did not disclose ethnicity or identified as other ethnicity; 81% ($n = 113$) had at least some college education, 15% ($n = 21$) were high school graduates, 1% ($n = 2$) did not graduate from high school, and 2% ($n = 3$) did not disclose education.

Stimuli

Providing a checklist is a commonly used short-term media literacy intervention tool and has been found to be effective (Ghenai and Mejova, 2018). One comparison condition was the NLM checklist.[1] The NLM checklist was introduced before the participants started to read and evaluate the information. To ensure the accessibility of the checklist, a screenshot was provided to participants. Then, before

---

1 https://medlineplus.gov/webeval/webevalchecklist.html

they read each of the four pieces of information, they were prompted to refer to the NLM checklist to assist in information evaluation.

The ChatGPT tool condition was manipulated by showing the participants ostensible screenshots of analysis of persuasive strategies in the information encountered (Figure 1). Participants were explained that ChatGPT is an AI tool that can be trained to dissect persuasive strategies to increase people's critical thinking for information evaluation. The response of ChatGPT of explaining the persuasive strategies in each piece of information was displayed right after they were exposed to the information. The control group was not provided with any media literacy tool.

In the pilot study, 79 participants were recruited from CloudResearch to complete the evaluation of six pieces of randomly displayed health-related information (Table 1). The articles were written with 9[2] of the 12 persuasive strategies embedded. Each article was sent into ChatGPT 3.5 and ChatGPT 3.5 was then asked to identify persuasive strategies in it. The authors modified the responses by ChatGPT 3.5 by removing errors and creating the stimuli of the ChatGPT condition.

By nature, some information is more difficult for people to accurately assess veracity, partly due to the novelty of the information.

---

2 Nine of the 12 strategies were features in the articles: fabricating misinformation via vivid storytelling; using personal experience or anecdotes rather than scientific findings as evidence; discrediting government or pharmaceutical companies; attacking science by exploiting its innate limitation; inappropriately using scientific evidence to support a false claim; making a conclusion based on biased reasoning; using fear or anger appeals in persuasion; using certain linguistic intensifiers to highlight the points; and establishing the legitimacy of false claims by using certain cues to activate people's credibility heuristics (e.g., medical jargon, seemingly credible source).

Two pieces of information were not selected in the main study because it was either too easy or too difficult to discern veracity. Four pieces were chosen for the main study, and they differed in terms of information evaluation difficulty as demonstrated by accuracy rate. Therefore, information evaluation difficulty was included as an independent variable in the main study.

Measures

Accuracy in information credibility evaluation

Participants rated the information credibility of each piece by indicating their agreement using a 5-point scale to rate the piped claim of the article as "believable," "authentic," and "accurate" (Appelman et al., 2016). Because participants read both true and misinformation, simply comparing the perceived message credibility would not capture accuracy in information credibility assessment. Therefore, we calculated the distance to ground truth, i.e., 5 being the ground truth for true information and 1 being the ground truth for misinformation. The greater the distance to ground truth, the less accuracy in information credibility evaluation.

Perceived usefulness of media literacy tool

Four items adapted from Taylor and Todd's (1995) scale were used to assess the perceived usefulness of the NLM checklist and the ChatGPT tool. Example items were as follows: "I think [piped tool] is useful for information evaluation" and "Overall, using [piped tool] for information evaluation will be advantageous."

Behavioral intention of future use

Three items were adapted from Davis' (1989) scale to assess future use intention. An example item was as follows: "Using

TABLE 1 Six pieces of information in the pilot study.

	Accuracy rate (%)	Main study
Misinformation		
Cashews can replace prescription drugs for treating depression	44	Yes
Preservatives in sunscreen cause breast cancer	39	No
Imported pet food might be contaminated with radioactive substances	60	Yes
True information		
Potential link between caffeine and reduced risk of Alzheimer's	73	Yes
Ginger can improve digestive health and relieve menstrual cramps	94	No
Lone star tick bite can make people allergic to red meat	50	Yes

[piped tool] for information evaluation is something I would do in the future."

All the above measures were rated on a 5-point Likert scale. Additionally, issue importance (Paek et al., 2012) and general information literacy (van der Vaart and Drossaert, 2017) were included as control variables.

Data analysis

To answer R1a, we conducted a mixed-model analysis of covariance (ANCOVA) to compare the effectiveness of the ChatGPT tool with the control, with information type (true information vs. misinformation) and information evaluation difficulty (hard vs. easy) as within-subjects factors, controlling for issue importance. A similar mixed-model ANCOVA was used to compare the ChatGPT tool with the benchmark of the NLM checklist, to answer RQ1b.[3] For RQ2 and RQ3, we did one-factor ANCOVA, controlling for information literacy to compare the ChatGPT tool with the NLM checklist in terms of perceived usefulness and future use intention of media literacy tools.

---

3 We also did a 3 × 2 × 2 ANCOVA and the results were comparable: there was a main effect of information type $F(1,542) = 8.35$, $p = 0.004$. An interaction effect was found between media literacy tools and information type $F(1,542) = 8.02$, $p < 0.001$. Issue importance was a significant covariate, $F(1,542) = 5.11$, $p = 0.024$. Similarly, a 2 × 2 × 2 ANCOVA was conducted to compare the NLM checklist with the control group. No main effects or two-way interactions were found. A marginally significant three-way interaction was found, $F(1,422) = 3.89$, $p = 0.049$.

Results

ChatGPT tool vs. control

To compare the ChatGPT tool with the control group (RQ1a), the ANCOVA results indicated a significant main effect of information type, $F(1,322) = 7.03$, $p = 0.008$, suggesting better accuracy in credibility evaluation in true information than misinformation. Additionally, a significant interaction effect emerged between media literacy tools and information type, $F(1,322) = 17.16$, $p < 0.001$. There was also a marginally significant interaction effect between information type and information evaluation difficulty, $F(1,322) = 3.93$, $p = 0.048$. No other effects were found significant. A subsequent *post-hoc* analysis focused on the interaction between media literacy tools and information type given its robustness. The *post-hoc* pairwise comparison with Tukey's adjustment revealed worse accuracy in information credibility evaluation in the ChatGPT tool condition only for misinformation, exhibiting a greater distance to ground truth ($M = 2.18$, $SE = 0.12$) than the control condition ($M = 1.68$, $SE = 0.10$), $p < 0.001$. Moreover, within the ChatGPT tool condition, accuracy in information credibility evaluation was markedly worse for misinformation ($M = 2.18$, $SE = 0.12$) than for true information ($M = 1.44$, $SE = 0.12$), $p < 0.001$, demonstrated by a greater distance to ground truth.

ChatGPT tool vs. NLM checklist

To answer RQ1b, the ANCOVA test revealed a significant main effect for information type, $F(1,338) = 16.45$, $p < 0.001$. Issue importance was also found to be a significant covariate, $F(1,338) = 8.70$, $p = 0.003$. An interaction effect was found between media literacy tools and information type, $F(1,338) = 6.66$, $p = 0.01$. No other effects were significant. The *post-hoc* pairwise comparison with Tukey's adjustment indicated that the ChatGPT tool resulted in less accuracy in information credibility evaluation of misinformation ($M = 2.19$, $SE = 0.13$) than true information ($M = 1.43$, $SE = 0.13$), $p < 0.001$, demonstrated by a greater distance to ground truth. However, no discernible differences were noted in the direct comparison between the NLM checklist and ChatGPT tool across both true ($p = 0.419$) and false misinformation ($p = 0.132$).

To answer RQ2, the one-factor ANCOVA demonstrated that there were no differences in the perceived usefulness of the ChatGPT tool ($M = 3.85$, $SD = 0.68$) and the NLM checklist ($M = 4.08$, $SD = 0.81$), $F(1,84) = 1.85$, $p = 0.177$. No statistically significant differences in future use intentions were found between the ChatGPT tool ($M = 3.19$, $SD = 1.08$) and the NLM checklist ($M = 3.65$, $SD = 1.04$), $F(1,84) = 3.91$, $p = 0.051$, answering RQ3. Note that general information literacy was a significant covariate, meaning that general information literacy was positively associated with the perception of the usefulness of media literacy tools, $F(1,84) = 7.71$, $p = 0.007$, as well as people's intention to use media literacy tools in the future, $F(1,84) = 4.83$, $p = 0.031$.

Discussion

Using dissecting persuasive strategies of ChatGPT can be equally effective when compared with the NLM checklist, and that

information type was a significant moderator such that ChatGPT was more effective in helping people identify true information than misinformation. However, using ChatGPT was worse than the control when it comes to misinformation discernment. The ChatGPT tool was not evaluated to be different from the NLM checklist in terms of usefulness and future use intention.

Despite that existing literature has documented that using chatbot-based approaches for combating misinformation could be feasible (Altay et al., 2022; Musi et al., 2023; Chan et al., 2017), our finding suggests that reading the rationale from ChatGPT for information veracity evaluation actually resulted in worse misinformation discernment compared to control. The findings of our study provided insights about several important factors to consider when it comes to testing the effectiveness of using chatbots as a media literacy tool. The first factor, which is also an important contribution of our current study, is the extent to which the studies mirror the complex information environment by mixing different information types (i.e., true and misinformation) and difficult levels (i.e., easy and hard) when asking participants to discern information credibility. Recent studies, such as the *Fake New Immunity Chatbot* (Musi et al., 2023), have reported the effectiveness of using chatbots to teach participants to recognize fallacies through reason-checking, but those studies were conducted in a context where participants learned persuasive strategies by analyzing misinformation only. In our study, both true and misinformation were presented to participants, which may have created more challenging tasks about information discernment. Our finding implies that future research testing media literacy tools for information discernment should include both misinformation and true information to better demonstrate the effectiveness of the tool in the complex information environment.

The second factor that may affect the effectiveness of a chatbot tool is related to interactivity and participants' engagement in learning. Previous literature has demonstrated that the game-based interventions were particularly successful due to their ability to interactively teach persuasive strategies (Buts, 2020). Our study involved participants' reading of pre-generated analysis of persuasive strategies from ChatGPT. The reading of texts is at a different level of interactivity and user engagement enabled by game-based interventions. Even though we have ensured that participants read and comprehended the analysis of persuasive strategies, it is possible that participants could not cognitively internalize the knowledge and skill points to connect persuasive strategies to the quality of information. For more complicated cognitive skills, such as learning persuasive strategies, it may be critical to create an enactive experience scaffolded by interactive steps. To better test the effectiveness of AI tools facilitating information discernment by dissecting persuasive strategies, future research should allow the participants to use the tools directly to gauge effectiveness.

Moreover, our findings showed that information literacy was positively associated with the perception of the usefulness of media literacy tools and their intention to use the media literacy tools in the future, although the perceived usefulness and use intention did not differ between the ChatGPT tool and the NLM checklist. This is consistent with the existing literature that has stressed the important role of information literacy in facilitating the identification of fake news (Jones-Jang et al., 2021). Our study reveals that people with higher levels of information literacy also appreciate media literacy tools more and are more likely to use them. This implies that such media literacy tools might benefit individuals who already have

somewhat sufficient information literacy. In other words, those who have low information literacy and mostly need assistance in information discernment may not take advantage of such tools, possibly resulting in "rich get richer." Future research may also explore how to encourage individuals with low information literacy to accept technological tools to facilitate their information discernment process.

The ChatGPT tool did not differ from the NLM checklist in terms of accuracy in information credibility assessment for both true and misinformation nor did it differ from the control for true information assessment. However, the ChatGPT tool was found to be more effective in helping people discern true information than misinformation. This was a promising finding because one of the concerns was that highlighting persuasive strategies in true information might heighten people's perception of persuasive intent, which may increase suspicion even for true information, resulting in distrust of credible information (Krause et al., 2022). The fact that the ChatGPT tool did not result in false positives—mistakenly identifying true information as false—was encouraging. In fact, extensive research findings on identifying misinformation do not easily translate into how people evaluate true information (Krause et al., 2022). Uncertainties and other sociopolitical factors associated with true information may make people dismiss it. The ChatGPT tool seems to have the potential to enhance people's confidence in verifying true information by providing reasoning.

An intriguing finding was that the ChatGPT tool performed worse than the control when it comes to misinformation. The lower effectiveness of the ChatGPT tool may be due to people's lack of trust in ChatGPT and psychological reactions triggered by viewing information analysis of ChatGPT. For example, people may have ethical concerns such as moral obligations and duties of AI and its creators (Siau et al., 2020). Biased or inappropriate content may still appear due to limitations related to the algorithms and training data (Zhou et al., 2023). People may feel their freedom and autonomy in assessing information be threatened when ChatGPT actively offers advice (Pizzi et al., 2021). Future research should examine the mediation mechanism, including but not limited to trust in ChatGPT and psychological reactions, to explain this unexpected effect.

Limitations

There are several limitations in our study. The first limitation is related to the combination of skill learning and outcome testing into one process. Future research may first establish the learning outcome; that is, people truly understand and acknowledge the usefulness of using persuasive strategies to identify misinformation and then presenting people with a different set of information to apply the knowledge and skills learned for information discernment. In this way, we could be more confident in disentangling the effects of using persuasive strategies as an approach to improve media literacy and using ChatGPT as a tool for dissecting persuasive strategies for people while they are processing information.

In addition, the ChatGPT condition had a lower number of participants, due to a larger number of participants failing the comprehension check. The fact that they failed the comprehension check showed that they did not understand the persuasive strategies dissected by the ChatGPT tool. This further demonstrated that it is important to first establish that participants understand the persuasive strategies more in-depth before they use the ChatGPT tool to dissect

the persuasive strategies just in time. Future research could investigate how to better present persuasive strategies to the users. For example, future research may inform participants about the ground truth in the first step of learning persuasive strategies so that they could establish the connection between the persuasive strategies and the veracity of information (i.e., the learning process) and then apply the knowledge and skills to information discernment tasks (i.e., the outcome testing). Finally, the ChatGPT condition in terms of its responses regarding the veracity of the articles was created by removing errors in the actual responses of ChatGPT of identifying the persuasive strategies. In other words, the findings were based on the assumption that ChatGPT was able to accurately identify persuasive strategies. Given that performance of ChatGPT has limitations, future research may explore how people's trust in capabilities of ChatGPT may have played a role in their reliance on the media literacy information provided by ChatGPT.

Conclusion

The current study examined the effectiveness of using dissecting persuasive strategies of ChatGPT as a media literacy tool to assist people's evaluation of online health information. Information type was a significant moderator such that ChatGPT was more effective in helping people identify true information than misinformation. This finding suggests that dissecting persuasive strategies of ChatGPT is promising to enhance people's confidence in verifying true information by providing reasoning, which has important implications for proper evaluations of true information given the complex information environment wherein true claims are not always clear-cut. In addition, while using dissecting persuasive strategies of ChatGPT can be equally effective when compared with the NLM checklist, the ChatGPT tool performed worse than control in terms of helping people discern misinformation. These findings suggest an alternative delivery of a media literacy program may be needed, such as using conversational features of ChatGPT so that people can interact and engage more in the learning of persuasive strategies.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Author contributions

WP: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. JM: Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing. T-WL: Formal analysis, Investigation, Writing – original draft.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Altay, S., Schwartz, M., Hacquin, A. S., Allard, A., Blancke, S., and Mercier, H. (2022). Scaling up interactive argumentation by providing counterarguments with a chatbot. *Nat. Hum. Behav.* 6, 579–592. doi: 10.1038/s41562-021-01271-w

Appelman, A., and Sundar, S. S. (2016). Measuring message credibility: construction and validation of an exclusive scale. *J. Mass Commun. Q.* 93, 59–79. doi: 10.1177/1077699015606057

Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., and van der Linden, S. (2021). Towards psychological herd immunity: cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data Soc.* 8, 1–18. doi: 10.1177/20539517211013868

Buts, J. (2020). Memes of Gandhi and mercury in anti-vaccination discourse. *Media Commun.* 8, 353–363. doi: 10.17645/mac.v8i2.2852

Chan, M. P. S., Jones, C. R., Jamieson, K. H., and Albarracín, D. (2017). Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol. Sci.* 28, 1531–1546. doi: 10.1177/0956797617714579

Choudhury, A., and Shamszare, H. (2023). Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis. *J. Med. Internet Res.* 25:e47184. doi: 10.2196/47184

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13, 319–340. doi: 10.2307/249008

De Vito, M. A., Gergle, D., and Birnholtz, J., editors. (2017) "Algorithms ruin everything": #RIPTwitter, folk theories, and resistance to algorithmic change in social media. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI); 2017 May 06–11; Denver, CO.

DeDominicis, K., Buttenheim, A. M., Howa, A. C., Delamater, P. L., Salmon, D., Omer, S. B., et al. (2020). Shouting at each other into the void: a linguistic network analysis of vaccine hesitance and support in online discourse regarding California law SB277. *Soc. Sci. Med.* 266:113216. doi: 10.1016/j.socscimed.2020.113216

Flaxman, S., Goel, S., and Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opin. Q.* 80:298. doi: 10.1093/poq/nfw006

Gallagher, J., and Lawrence, H. Y. (2020). Rhetorical appeals and tactics in New York times comments about vaccines: qualitative analysis. *J. Med. Internet Res.* 22:e19504. doi: 10.2196/19504

Ghenai, A., and Mejova, Y. (2018). Fake cures: user-centric modeling of health misinformation in social media. *Proc. ACM Hum-Comput. Interact.* 2, 1–20. doi: 10.1145/3274327

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., et al. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proc. Natl. Acad. Sci. USA* 117, 15536–15545. doi: 10.1073/pnas.1920498117

Guo, Z. J., Schlichtkrull, M., and Vlachos, A. (2022). A survey on automated fact-checking. *Trans. Assoc. Comput. Ling.* 10:178. doi: 10.1162/tacl_a_00454

Haupt, M. R., Li, J., and Mackey, T. K. (2021). Identifying and characterizing scientific authority-related misinformation discourse about hydroxychloroquine on twitter using unsupervised machine learning. *Big Data Soc.* 8:20539517211013843. doi: 10.1177/20539517211013843

Hobbs, R. (2010) Digital and media literacy: A plan of action. A white paper on the digital and media literacy recommendations of the knight commission on the information needs of communities in a democracy: ERIC.

Jones-Jang, S. M., Mortensen, T., and Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *Am. Behav. Sci.* 65, 371–388. doi: 10.1177/0002764219869406

Kearney, M. D., Selvan, P., Hauer, M. K., Leader, A. E., and Massey, P. M. (2019). Characterizing HPV vaccine sentiments and content on Instagram. *Health Educ. Behav.* 46, 37–48. doi: 10.1177/1090198119859412

Kou, Y., Gui, X., Chen, Y., and Pine, K. (2017). Conspiracy talk on social media: collective sensemaking during a public health crisis. *Proc. ACM Hum-Comput. Interact.* 1, 1–21. doi: 10.1145/3134696

Krause, N. M., Freiling, I., and Scheufele, D. A. (2022). The "infodemic" infodemic: toward a more nuanced understanding of truth-claims and the need for (not) combatting misinformation. *Ann. Am. Acad. Pol. Soc. Sci.* 700, 112–123. doi: 10.1177/00027162221086263

Lu, C., Hu, B., Bao, M. M., Wang, C., Bi, C., and Ju, X. D. (2024). Can media literacy intervention improve fake news credibility assessment? A meta-analysis. *Cyberpsychol. Behav. Soc. Networking* 27, 240–252. doi: 10.1089/cyber.2023.0324

Mingoia, J., Hutchinson, A. D., Gleaves, D. H., and Wilson, C. (2019). The impact of a social media literacy intervention on positive attitudes to tanning: a pilot study. *Comput. Human Behav.* 90:188. doi: 10.1016/j.chb.2018.09.004

Musi, E., Carmi, E., Reed, C., Yates, S., and O'Halloran, K. (2023). Developing misinformation immunity: how to reason-check fallacious news in a human-computer interaction environment. *Soc. Media Soc.* 9, 1–18. doi: 10.1177/20563051221150407

Nan, X. L., Wang, Y., and Thier, K. (2022). Why do people believe health misinformation and who is at risk? A systematic review of individual differences in susceptibility to health misinformation. *Soc. Sci. Med.* 314:115398. doi: 10.1016/j.socscimed.2022.115398

O'Shaughnessy, M. R., Schiff, D. S., Varshney, L. R., Rozell, C. J., and Davenport, M. A. (2023). What governs attitudes toward artificial intelligence adoption and governance? *Sci. Public Policy* 50, 161–176. doi: 10.1093/scipol/scac056

Paek, H. J., Hove, T., Kim, M., Jeong, H. J., and Dillard, J. P. (2012). When distant others matter more: perceived effectiveness for self and other in the child abuse PSA context. *Media Psychol.* 15, 148–174. doi: 10.1080/15213269.2011.653002

Peng, W., Lim, S., and Meng, J. B. (2023). Persuasive strategies in online health misinformation: a systematic review. *Inf. Commun. Soc.* 26, 2131–2148. doi: 10.1080/1369118x.2022.2085615

Pennycook, G., Bear, A., Collins, E. T., and Rand, D. G. (2020). The implied truth effect: attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Manag. Sci.* 66, 4944–4957. doi: 10.1287/mnsc.2019.3478

Pizzi, G., Scarpi, D., and Pantano, E. (2021). Artificial intelligence and the new forms of interaction: who has the control when interacting with a chatbot? *J. Bus. Res.* 129:878. doi: 10.1016/j.jbusres.2020.11.006

Prasad, A. (2022). Anti-science misinformation and conspiracies: COVID–19, post-truth, and science & technology studies (STS). *Sci. Technol. Soc.* 27, 88–112. doi: 10.1177/09717218211003413

Qian, S. J., Shen, C. H., and Zhang, J. W. (2023). Fighting cheapfakes: using a digital media literacy intervention to motivate reverse search of out-of-context visual misinformation. *J. Comput.-Mediat. Commun.* 28, 1–12. doi: 10.1093/jcmc/zmac024

Reynolds-Tylus, T. (2019). Psychological reactance and persuasive health communication: a review of the literature. *Front. Commun.* 4. doi: 10.3389/fcomm.2019.00056

Rheu, M., Shin, J. Y., Peng, W., and Huh-Yoo, J. (2021). Systematic review: trust-building factors and implications for conversational agent design. *Int. J. Hum. Comput. Interact.* 37, 81–96. doi: 10.1080/10447318.2020.1807710

Romer, D., and Jamieson, K. H. (2020). Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. *Soc. Sci. Med.* 263:113356. doi: 10.1016/j.socscimed.2020.113356

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., et al. (2020). Susceptibility to misinformation about COVID-19 around the world. *R. Soc. Open Sci.* 7, 1–15. doi: 10.1098/rsos.201199

Roozenbeek, J., and van der Linden, S. (2019). The fake news game: actively inoculating against the risk of misinformation. *J. Risk Res.* 22, 570–580. doi: 10.1080/13669877.2018.1443491

Salvador Casara, B. G., Suitner, C., and Bettinsoli, M. L. (2019). Viral suspicions: vaccine hesitancy in the web 2.0. *J. Exp. Psychol. Appl.* 25, 354–371. doi: 10.1037/xap0000211

Siau, K., and Wang, W. Y. (2020). Artificial intelligence (AI) ethics: ethics of AI and ethical AI. *J. Database Manag.* 31, 74–87. doi: 10.4018/jdm.2020040105

Taylor, S., and Todd, P. A. (1995). Understanding information technology usage: a test of competing models. *Inf. Syst. Res.* 6, 144–176. doi: 10.1287/isre.6.2.144

van der Vaart, R., and Drossaert, C. (2017). Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. *J. Med. Internet Res.* 19:e27. doi: 10.2196/jmir.6709

Vraga, E. K., Bode, L., and Tully, M. (2022). Creating news literacy messages to enhance expert corrections of misinformation on twitter. *Commun. Res.* 49, 245–267. doi: 10.1177/0093650219898094

Walter, N., Brooks, J. J., Saucier, C. J., and Suresh, S. (2021). Evaluating the impact of attempts to correct health misinformation on social media: a meta-analysis. *Health Commun.* 36, 1776–1784. doi: 10.1080/10410236.2020.1794553

Zarocostas, J. (2020). How to fight an infodemic. *Lancet* 395:676. doi: 10.1016/S0140-6736(20)30461-X

Zhan, E. S., Molina, M. D., Rheu, M., and Peng, W. (2023). What is there to fear? Understanding multi-dimensional fear of AI from a technological affordance perspective. *Int. J. Hum. Comput. Interact.*, 1–18. doi: 10.1080/10447318.2023.2261731

Zhang, J. W., Featherstone, J. D., Calabrese, C., and Wojcieszak, M. (2021). Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. *Prev. Med.* 145:106408. doi: 10.1016/j.ypmed.2020.106408

Zhang, L. L., Iyendo, T. O., Apuke, O. D., and Gever, C. V. (2022). Experimenting the effect of using visual multimedia intervention to inculcate social media literacy skills to tackle fake news. *J. Inf. Sci.* doi: 10.1177/016555152211317

Zhao, X. L., Chen, L., Jin, Y. C., and Zhang, X. Z. (2023). Comparing button-based chatbots with webpages for presenting fact-checking results: a case study of health information. *Inf. Process. Manag.* 60:103203. doi: 10.1016/j.ipm.2022.103203

Zhou, J. W., Zhang, Y. X., Luo, Q. N., Parker, A. G., and De Choudhury, M. editors. (2023) Synthetic lies: understanding AI-generated misinformation and evaluating algorithmic and human solutions. CHI conference on human factors in computing systems (CHI), April 23–28, 2023; Hamburg, Germany.