scientific reports

Check for updates

OPEN

Conditional effects of source expertise and pre-existing attitudes on objective knowledge in AI-assisted health information verification

Jihyeon Oh[1,2], Christian Montag[3,4,5], Julian Kohne[6] & Peter Johannes Schulz[1,2,7,8]✉

Generative AI systems are increasingly integrated into health communication, yet it remains unclear under what conditions source-related cues shape objective knowledge outcomes when individuals verify health information with AI assistance. This study examines the roles of source expertise, pre-existing attitudes, and knowledge states in shaping objective knowledge during AI-assisted health information verification. In an experiment with 103 participants, individuals viewed a mixed-accuracy Facebook post about gluten-free diets attributed to either an expert-labeled or a non-expert-labeled source, then used ChatGPT to verify the information. Source expertise alone did not enhance objective knowledge, but its effect emerged among participants with favorable pre-existing attitudes. The predicted moderating role of knowledge state (uncertain vs. misinformed) was not supported, although exploratory patterns indicated greater responsiveness among uncertain users. The findings suggest that dual-process perspectives are informative for understanding AI-assisted information processing in contexts where verification is supported by generative AI.

Keywords Generative AI, Source expertise, Pre-existing attitude, Knowledge state, Objective knowledge

The growing use of conversational artificial intelligence (AI) is reshaping how individuals access and make sense of health information. Instead of independently navigating multiple sources, users increasingly rely on AI systems such as ChatGPT to locate and summarize online content[1,2]. Automated systems now assist users in searching, organizing, and interpreting content, which shifts part of the cognitive process from individuals to technology (Grinschgl & Neubauer, 2022). In such settings, AI-supported information processing can be understood as a form of distributed cognition. During this process, people offload parts of the interpretive task to AI systems[3] while remaining alert to the reliability of the responses[4]. Rather than evaluating accuracy entirely on their own, users work with AI to organize and interpret information, maintaining a balance between delegation and oversight[5,6]. This dynamic suggests that AI-assisted verification involves distinct cognitive processes that may shape how people process and interpret information. Accordingly, in this study, AI systems such as ChatGPT are conceptualized as verification environments rather than independent sources of influence. Rather than being manipulated as an independent variable, ChatGPT served as a shared verification affordance, allowing the study to examine how classic credibility and motivational processes operate when AI-assisted verification is available.

This process potentially also plays a role for information about gluten-free diets. Initially prescribed for individuals with celiac disease and related conditions, gluten-free diets have entered mainstream discourse, shaped by commercial promotion and popular beliefs[7]. In such mixed environments where scientific and lay

[1]Division of Communication & Media, Ewha Womans University, Seoul, Republic of Korea. [2]Institute of Communication & Health, Faculty of Communication, Culture and Society, University of Lugano, Lugano, Switzerland. [3]Centre for Cognitive and Brain Sciences, Institute of Collaborative Innovation, University of Macau, Macao, SAR, China. [4]Faculty of Science and Technology, University of Macau, Macao SAR, China. [5]Faculty of Social Sciences, University of Macau, Macao SAR, China. [6]Department of Computational Social Science, GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany. [7]Wee Kim Wee School of Communication and Information/LKC School of Medicine, Nanyang Technological University, Singapore, Singapore. [8]Institute of Communication & Health, Faculty of Communication, Culture and Society, University of Lugano, Via G. Buffi 13, Lugano 6900, Switzerland. ✉email: schulzp@usi.ch

narratives intersect, source expertise can serve as an anchor for evaluation. At the same time, pre-existing attitudes influence how individuals interpret and accept information.

Prior research has emphasized the role of source expertise and prior attitudes in shaping information evaluation. Dual-process models such as the elaboration likelihood model[8] and the heuristic–systematic model[9] explain that expertise signals reliability and guides evaluation either heuristically or through elaboration. Empirical studies have shown that expert sources enhance perceived credibility and influence evaluative outcomes[10–12]. More recent studies show that expertise and attitudes can interact, such that individuals with firm beliefs interpret identical information differently depending on the perceived credibility of the source[13–16]. Yet it remains unclear how these relationships apply when information is processed with the assistance of AI systems.

Another factor concerns individuals' knowledge state. Recent research highlights the importance of differentiating between two forms of knowledge failure: uncertainty (being uninformed and recognizing a lack of knowledge) and misinformation (holding false beliefs confidently)[17,18]. Although both can lead to errors, they have distinct implications for health behavior and communication responses[19,20]. Uncertain individuals are often epistemically open, more likely to seek further information, and more responsive to corrective input[21,22]. In contrast, misinformed individuals resist disconfirming evidence through motivated reasoning, and higher subjective confidence in their (incorrect) knowledge can even reinforce harmful behaviors. Testing this distinction is crucial, because treating all knowledge failures as equivalent overlooks the unique risks posed by the misinformed.

Building on these observations, the present study examines how individuals interpret and learn from mixed-accuracy information about gluten-free diets when assisted by ChatGPT. Specifically, it investigates (1) the main effect of source expertise, (2) the moderating role of pre-existing attitudes, and (3) the conditional influence of knowledge state. To model this process, the experimental procedure combined exposure to a Facebook post with an AI-assisted search using ChatGPT, simulating how users encounter and verify mixed-accuracy information online. This AI-assisted verification setting provides a context in which expertise cues, motivational factors, and knowledge states jointly shape how individuals evaluate and learn from mixed-accuracy information. This design represents a mixed-accuracy digital learning context rather than a persuasion setting, focusing on how users acquire and evaluate information with AI assistance.

Literature review

Source expertise and objective knowledge acquisition

Source expertise serves as a key heuristic for judging information credibility. Both the elaboration likelihood model[8] and the heuristic–systematic model[9] explain that expertise signals reliability and guides how recipients evaluate information. Expert sources enhance perceived credibility[11] and strengthen persuasive impact relative to non-expert sources[10,12]. In digital environments, expertise also functions as a salient heuristic that facilitates judgments under conditions of information overload[23]; Sundar, 2008).

Prior studies, however, have primarily focused on evaluative outcomes such as credibility judgments[10], attitude change[12], and perceived truth[24]. For instance, Wertgen and Richter[25], showed that source credibility influenced readers' evaluation of implausible claims, and Nadarevic et al.[24], found that statements attributed to highly credible sources were judged as more truthful. A recent review further concluded that effects of source credibility are generally stronger for evaluative judgments than for knowledge-related outcomes[26]. Overall, these findings suggest that expertise reliably shapes judgments of trust, plausibility, and truth, yet it remains unclear whether such effects extend to objective knowledge outcomes. Some evidence further suggests that source expertise may influence how information is processed, beyond its role in evaluative judgment. For example, text comprehension research shows that highly credible sources promote more intensive, meaning-level validation of message information[25], a processing orientation that is considered conducive to the integration and updating of factual knowledge. Such engagement may facilitate knowledge acquisition by promoting the integration of factual content from expert sources.

Beyond evaluative judgments, some research has examined how expert sources influence factual accuracy in corrective contexts. For example, corrections from expert or official sources have been shown to effectively reduce misperceptions about health issues[27]. Similarly, Wang[16], demonstrated that corrective messages from expert sources were more effective in reducing misperceptions about genetically modified food safety than those from non-expert sources. However, these studies focus on belief correction rather than on changes in objectively assessed factual knowledge measured using performance-based tests. Taken together, prior work highlights the evaluative role of expertise but provides only limited and conditional evidence regarding its impact on objective knowledge outcomes. Against this background, the present study examines whether expert-labeled sources are associated with higher levels of objective knowledge than non-expert-labeled sources.

H1. Participants exposed to an expert-labeled source will demonstrate higher levels of objective knowledge than participants exposed to a non-expert-labeled source.

Pre-existing attitude as a motivational moderator

Dual-process theories emphasize that motivation determines how deeply individuals process message content and source cues. According to the elaboration likelihood model, individuals who are more motivated or personally invested in an issue engage in more systematic processing[8]. Prior research has shown that people with strong pre-existing attitudes process attitude-relevant information more elaborately[12,28], and that deeply processed attitudes tend to be more persistent and resistant to change[29].

Empirical findings support this conditional view. Individuals with stronger prior attitudes show heightened sensitivity to credibility cues that are consistent with their beliefs[13–16]. For instance, Schweiger and Cress[14], found that source credibility more strongly influenced information selection among participants with higher attitude confidence. Similarly, Wang[16], demonstrated that expert-source corrections were more effective

when participants' prior beliefs aligned with the corrective message, and Liu et al.[13], observed that prior belief alignment moderated fact-checking effectiveness. Together, these findings suggest that the persuasive impact of source expertise is contingent on recipients' motivational engagement with the topic.

However, most prior work has examined these contingencies for attitude change and belief correction rather than factual learning. Research on misinformation correction shows that motivation, attitudes, and source trust predict whether corrections are accepted[30], yet it remains unclear whether such motivational effects extend to objective knowledge outcomes. Prior studies nonetheless indicate that favorable pre-existing attitudes are associated with greater elaboration and responsiveness to credibility cues[14,28]. Favorable attitudes may therefore be linked to greater engagement with expert-provided information, with potential implications for how factual content is processed and evaluated. Such engagement may support the integration of factual content from expert sources into objective knowledge. Therefore, the present study tests whether the positive effect of source expertise on objective knowledge is stronger among participants with favorable pre-existing attitudes toward gluten-free diets.

H2. The positive effect of source expertise on objective knowledge will be stronger among participants with more favorable pre-existing attitudes toward gluten-free diets.

Knowledge state as a boundary condition

Knowledge state distinguishes between two qualitatively different conditions: uncertainty and misinformation[17,18]. Misinformed individuals confidently hold false beliefs and resist correction[19,20], whereas uncertain individuals recognize gaps in their understanding and remain receptive to credible cues[21,22]. Empirical research further shows that misinformation can persist even after correction[30], while corrective messages are more effective when they align with existing worldviews[31]. Taken together, prior work suggests that uncertainty and misinformation differ not only in belief accuracy but also in the degree of confidence with which beliefs are held, a distinction relevant to how individuals respond to new information.

The distinction between uncertainty and misinformation is particularly important in the context of gluten-free diets. Some individuals are uncertain, recognizing gaps in their knowledge (e.g., "I am not sure whether gluten-free diets are healthier for people without celiac disease"). Others are misinformed, holding confidently to inaccurate beliefs (e.g., "Gluten-free diets are always healthier for everyone"). When individuals are uncertain and lack firm belief commitments, they may be more likely to treat new information as diagnostically useful and to rely on external cues such as source expertise when interpreting and evaluating new information. In contrast, misinformed individuals tend to hold inaccurate beliefs with greater confidence, which is associated with a higher likelihood of defensive processing and reduced receptivity to corrective information.

Building on this distinction, the effects of expertise cues are expected to differ between the two groups. Uncertain individuals may be more likely to rely on source expertise when interpreting and evaluating new information, whereas misinformed individuals are likely to discount such cues to protect existing beliefs. Accordingly, uncertainty is more likely to be accompanied by greater reliance on credibility cues when interpreting new claims, whereas misinformation is more likely to be accompanied by defensive processing that limits updating. As a result, the interaction between source expertise and pre-existing attitudes should emerge primarily among uncertain individuals but not among those who are misinformed.

Although knowledge state and prior attitude may appear related, prior work conceptualizes them as distinct constructs. Attitudes are motivational orientations that guide evaluation and behavioral intentions[28,32,33], whereas knowledge states concern the accuracy and confidence of factual beliefs[18]. This distinction is important because differences in belief confidence and epistemic stance are associated with variation in how individuals respond to expert cues, beyond differences in evaluative orientation toward the topic. Taken together, these insights suggest that knowledge state serves as a boundary condition moderating the joint influence of source expertise and prior attitudes.

H3. The interaction between source expertise and pre-existing attitudes will be more pronounced among uncertain participants than among misinformed participants.

Methods

Participants and procedure

A total of 103 participants took part in the study and were randomly assigned to one of two experimental conditions: a non-expert-labeled source condition ($n = 43$) or an expert-labeled source condition ($n = 60$). The average age was 21.86 years (SD = 7.52) in the non-expert-labeled condition and 20.95 years (SD = 6.52) in the expert-labeled condition. Gender distribution was comparable across conditions, with 30.2% male participants in the non-expert-labeled condition and 15.0% in the expert-labeled condition. Educational background was also similar, with most participants in both conditions enrolled in undergraduate programs (non-expert-labeled: M = 2.14, SD = 0.47; expert-labeled: M = 2.10, SD = 0.35).

In terms of language fluency, nearly all participants identified as fluent speakers of the survey language (non-expert-labeled: 100%; expert-labeled: 96.7%). Prior familiarity with gluten-related information was similarly low in both conditions (non-expert-labeled: M = 1.95, SD = 0.21; expert-labeled: M = 1.93, SD = 0.25), and previous experience with ChatGPT was likewise comparable (non-expert-labeled: M = 1.05, SD = 0.21; expert-labeled: M = 1.03, SD = 0.18).

Participants were randomly assigned to one of two experimental conditions: exposure to a Facebook post attributed to an expert-labeled source or to a non-expert-labeled source. Each post contained a mix of accurate and inaccurate information about gluten-free diets. After viewing the assigned post, participants were instructed to use ChatGPT to verify the accuracy of the information presented in the post by searching for additional information. This verification task was completed prior to answering the outcome measures. As part of the study procedure, participants were asked to paste a link to their ChatGPT conversation, thus linking their conversations

with ChatGPT to the survey responses and providing them to the researchers. Participants were not given specific instructions regarding how to formulate prompts or how many follow-up questions to ask. Prompting strategies and AI responses were not standardized, allowing individual-level factors to operate within the AI-assisted verification context. All study tasks were completed via an online survey platform. Before exposure, participants completed a pretest questionnaire assessing baseline self-assessed knowledge of gluten-free diets and pre-existing attitudes toward the topic. Following the ChatGPT interaction, they completed a posttest survey assessing objective knowledge, evaluation of the information search experience, trust in ChatGPT, and intention to share the information.

Preregistration

The study was preregistered on the Open Science Framework (OSF; *Evaluating ChatGPT's Effectiveness in Health-Related Information Retrieval*), registered on September 14, 2024.

Ethics statement

The study used fully anonymized data and did not involve any identifiable information. The Ethics Committee of Ulm University determined that the project is not subject to mandatory ethical review ("nicht beratungspflichtig"), and therefore no approval number was issued. All procedures adhered to relevant guidelines and regulations, and informed consent was obtained from all participants.

Stimulus materials

Choice of topic: Gluten-free diets

Gluten-free diets provide a theoretically suitable context for examining source expertise and knowledge processing in AI–mediated health communication. Unlike topics where expert consensus is strong and public debate is limited (e.g., cancer screening guidelines), gluten-free diets occupy a hybrid discursive space: they are medically essential for individuals with celiac disease but have also been popularized in mainstream culture through lifestyle promotion and commercial marketing. Prior research shows that many consumers adopt gluten-free diets without medical necessity based on the belief that they are generally healthier, despite limited evidence supporting such claims[34]. This duality generates a communication environment in which scientific evidence coexists with anecdotal claims, celebrity endorsements, and commercial persuasion. Such an environment is especially prone to mixed-accuracy information and thus creates the very conditions in which users must actively rely on source expertise cues to evaluate claims.

Moreover, gluten-free diets are uniquely positioned to reveal the influence of pre-existing attitudes and knowledge states. For some, they are viewed as medically necessary; for others, as a personal lifestyle choice; and for yet others, as unnecessary or even harmful. These divergent attitudes make gluten-free diets a fertile ground for testing how motivational filters shape the processing of source cues. Finally, the topic exemplifies the challenge of distinguishing uncertainty (e.g., "I don't know whether gluten-free diets are healthier") from misinformation (e.g., "gluten-free diets are healthier for everyone"), making it an ideal case to test the conceptual distinction between these cognitive states.

Message stimuli

Two Facebook posts were developed to manipulate an expert label attached to the source: one attributed to an expert-labeled source, "Food Network" (Fig. 1), and the other to a non-expert-labeled source, "Emma Kim" (Fig. 2). Both versions followed the same structure, with sections describing the causes and effects of gluten, the health risks of consumption, and dietary management recommendations. The content deliberately incorporated a mix of accurate statements (e.g., the association between gluten and celiac disease) and inaccurate claims (e.g., assertions that gluten causes diabetes or neurological disorders in the general population). This design reflects how online health discourse often presents accurate details alongside misleading or exaggerated assertions[35]. Presenting such mixed-accuracy content enables the study to evaluate not only how source expertise influences message processing, but also how effectively participants can distinguish and retain accurate information after interacting with ChatGPT. All message elements were held constant across conditions except for the attributed source identity, which served as the expertise cue via source labeling. Prior research has shown that such minimal yet salient manipulations can meaningfully shift information evaluation in digital environments[36].

Measures

Attitude toward gluten free diet

We adapted items from Sainsbury and Mullan[37], who measured attitudes toward maintaining a gluten-free diet using semantic differential scales grounded in the theory of planned behavior[32]. In our study, participants evaluated "being on a gluten-free diet" on three adapted 7-point semantic differential items: painful–not painful, unpleasant–pleasant, and burdensome–easy to follow. Higher scores indicated a more favorable attitude toward adopting a gluten-free diet (Cronbach's $\alpha = 0.80$).

Objective knowledge regarding gluten-free diet

As an objective, performance-based measure, a 15-item gluten knowledge test was used, with seven items (Q1–Q7) based on the experimental stimuli and eight items (Q8–Q15) adapted from Leffler et al.[38],. Each item was phrased as a factual assertion, and respondents indicated whether it was "true," "false," or "don't know." Among the 15 items, seven were classified as true statements (items 1, 3, 5, 6, 9, 10, and 13) and eight as false (items 2, 4, 7, 8, 11, 12, 14, and 15). The full list of items is provided in Appendix 1. The analyses used categorized groups based on the number of correct, incorrect, and "don't know" responses. On average, participants provided 9.57 (SD = 2.36) correct answers, 1.90 (SD = 1.72) incorrect answers, and 3.52 (SD = 2.26) "don't know" responses.

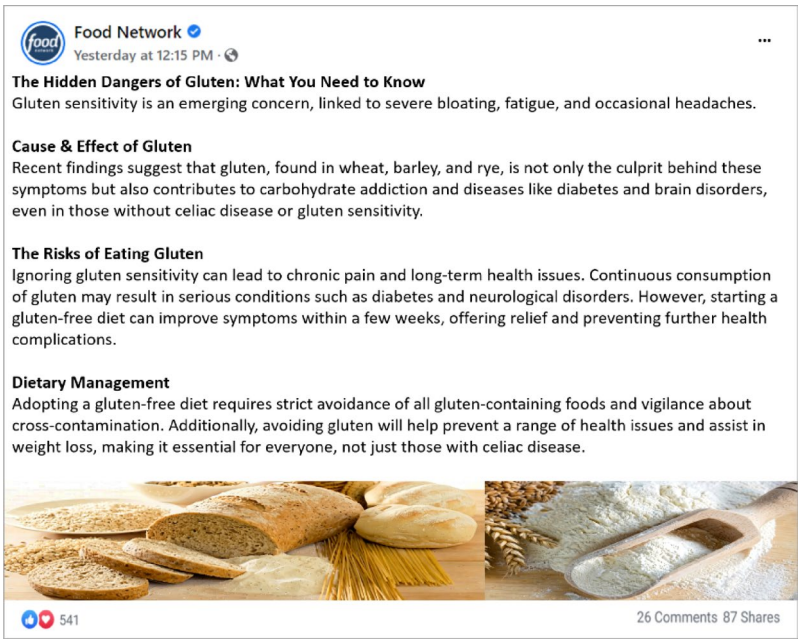


Fig. 1. Facebook post from expert-labeled source. The image was AI-generated by the authors using Stable Diffusion XL 1.0 (https://stablediffusionweb.com/).

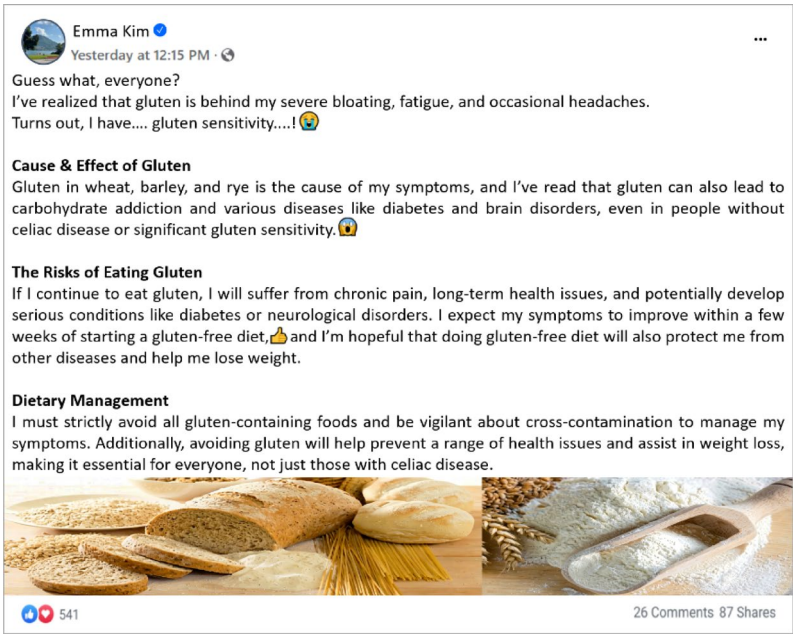


Fig. 2. Facebook post from non-expert-labeled source. The image was AI-generated by the authors using Stable Diffusion XL 1.0 (https://stablediffusionweb.com/).

A second objective knowledge indicator was constructed to reflect participants' relative orientation toward uncertainty versus misinformation. Incorrect responses (answering "false" to true items or "true" to a false one) and "don't know" responses were used to derive this indicator. Specifically, the number of incorrect responses was subtracted from the number of "don't know" responses. Higher values reflect a greater prevalence of uncertainty responses relative to incorrect responses, whereas lower values reflect a greater prevalence of incorrect responses. The resulting index ranged from −12 to +13 (M = 1.62, SD = 3.24). Accordingly, this indicator is intended to capture a task-specific epistemic orientation reflected in response patterns within the knowledge test, rather than a stable individual difference or enduring epistemic trait.

Pre-existing self-assessed knowledge
To measure pre-existing knowledge about gluten-free diets, participants completed three self-report items assessing their perceived knowledge and confidence prior to stimulus exposure. Using a 7-point scale (− 3 = very low/strongly disagree, 3 = very high/strongly agree), participants rated the following statements: "How would you rate your general knowledge of a gluten-free diet?", "If you had to identify foods that contain gluten, would you say you are confident in doing so?", and "I understand the dietary requirements of a gluten-free diet well." Responses were averaged to create a composite measure of self-assessed pre-existing knowledge, with higher scores indicating greater perceived knowledge (Cronbach's α = 0.90).

Evaluation of information search
To measure participants' evaluation of the information obtained through ChatGPT, we adapted items from Sachdeva's[39] consumer decision-making efficacy scale, particularly the information search dimension. Using a 7-point scale (− 3 = very low/strongly disagree, 3 = very high/strongly agree), participants rated the extent to which they agreed with three statements: "After the information search with the help of ChatGPT, I can assess my general knowledge about a gluten-free diet," "If I had to identify foods that contain gluten, I would feel confident," and "I understand the dietary requirements of a gluten-free diet well." The average was calculated and used as the evaluation of information search score (Cronbach's α = 0.74).

Trust in ChatGPT
To measure participants' trust in ChatGPT, we adapted an item from prior research[40,41]. Participants rated their agreement with the statement, "I trust ChatGPT from OpenAI," on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). Higher scores indicated greater trust in ChatGPT.

Statistical analysis
All analyses were conducted using SPSS Statistics Version 26 with the PROCESS macro version 3.5[42]. Hierarchical multiple regression was used to test H1, with objective knowledge (number of correct responses) as the dependent variable. Source expertise (expert-labeled = 1, non-expert-labeled = 0), pre-existing attitude toward gluten-free diets (mean-centered), and their interaction term were entered as predictors, while evaluation of the information search experience and trust in ChatGPT were included as covariates. PROCESS Model 1 was used to test H2, examining whether pre-existing attitude moderated the association between source expertise and objective knowledge. PROCESS Model 3 was used to test H3, assessing whether this interaction was further moderated by knowledge state (uncertain vs. misinformed). Finally, exploratory subgroup analyses were conducted separately for uncertain and misinformed participants to further probe the interaction patterns. All p-values are based on two-tailed tests. To assess whether the sample size was adequate for detecting interaction effects, a sensitivity power analysis was conducted using G*Power. With $N = 103$, $\alpha = 0.05$, and power set to 0.80, the analysis indicated sufficient sensitivity to detect small-to-moderate two-way interaction effects ($f^2 \approx 0.07$).

Results
Prior to analyzing the hypothesized model, descriptive statistics and bivariate correlations across all variables were computed. Table 1 presents the means, standard deviations, and correlations of all variables used in the study.

Main association analysis
To test H1, we conducted hierarchical multiple regression analyses with objective knowledge as the dependent variable. Table 2 summarizes the results of this regression analysis.

In Step 1, two covariates—evaluation of the information search and trust in ChatGPT—were entered into the model. Neither variable significantly predicted objective knowledge, and the model was not significant ($\Delta R^2 = 0.057$, $F(2, 100) = 2.997$, $p = .054$). In Step 2, source expertise was added to the model, but it was not a significant predictor of objective knowledge ($B = 0.63$, $p = .175$). The inclusion of source expertise led to a small, nonsignificant increase in explained variance ($\Delta R^2 = 0.017$, F change$(1, 99) = 1.87$, $p = .175$). In sum, the results indicate that source expertise did not significantly predict participants' objective knowledge after controlling for covariates. Thus, H1 was not supported.

	1	2	3	4	5	6
1. Evaluation of information search	1					
2. Trust in ChatGPT	0.17	1				
3. Source expertise	− 0.03	− 0.09	1			
4. Attitude toward gluten-free diet	− 0.02	0.12	0.05	1		
5. Pre-existing self-assessed knowledge	0.70**	0.05	0.03	0.06	1	
6. Objective Knowledge (posttest)	0.20*	0.16	− 0.15	− 0.01	0.14	1
M	4.40	3.23	0.42	2.78	3.42	9.57
SD	1.10	0.88	0.50	1.11	1.50	2.36

Table 1. Means, standard deviations, and correlations of all variables. *$p < .05$, **$p < .01$, ***$p < .001$.

	DV: Objective knowledge			
Step 1	B	SE	t	*p*
Evaluation of information search	0.39	0.21	1.84	0.069
Trust in ChatGPT	0.34	0.27	1.29	0.199
Model summary: $R^2 = 0.057$, $\Delta R^2 = 0.057$, $F(2, 100) = 2.997$, $p = .054$				
Step 2				
Evaluation of information search	0.39	0.21	1.83	0.070
Trust in ChatGPT	0.31	0.27	1.18	0.243
Source expertise	0.63	0.46	1.37	0.175
Model summary: $R^2 = 0.074$, $\Delta R^2 = 0.017$, $F(3, 99) = 2.637$, $p = .054$ (overall model); F change$(1, 99) = 1.866$, $p = .175$ (for added variable).				

Table 2. Results of Hierarchical Regression on Objective Knowledge.

Predictors	Objective knowledge	
	B	t
Source expertise	−2.368	−1.936
Pre-existing attitude	− 0.637	−2.060*
Source expertise X Pre-existing attitude	1.072	2.637**
R^2	0.136	
F	3.055*	

Table 3. Moderated effect: Pre-existing Attitude as a moderator in the association between Source Expertise and Objective Knowledge. *p<.05, **p<.01, ***p<.001.

As an additional analysis, we conducted an ANCOVA predicting objective knowledge while controlling for pretest self-assessed knowledge assessed prior to exposure. The effect of the expert-labeled condition remained non-significant, $F(1, 98) = 1.87$, $p = .174$, partial $\eta^2 = 0.019$.

Moderation analysis

To test H2, we conducted a moderation analysis using PROCESS Model 1 with objective knowledge as the dependent variable, source expertise as the independent variable, and pre-existing attitude as the moderator. Table 3 summarizes the results.

Source expertise did not have a significant main effect on objective knowledge ($B = -2.37$, $t = -1.94$, $p = .056$). This negative main effect should be interpreted in light of the significant interaction reported below, indicating that the effect of source expertise on objective knowledge is conditional on pre-existing attitudes and becomes positive only at higher levels of favorable attitudes. However, pre-existing attitude was a significant predictor ($B = -0.64$, $t = -2.06$, $p = .042$). Importantly, the interaction between source expertise and pre-existing attitude was statistically significant ($B = 1.07$, $t = 2.64$, $p = .010$, $\Delta R^2 = 0.062$, 95% CI [0.27, 1.88]; see Fig. 3). The full model explained 13.6% of the variance ($R^2 = 0.136$, $F(5, 97) = 3.05$, $p = .013$).

To further examine this interaction, conditional effects of source expertise were tested at low (−1 SD), mean, and high (+1 SD) levels of pre-existing attitude (Table 4). When pre-existing attitude was low, source expertise did not significantly predict objective knowledge (Effect = −0.56, Boot SE = 0.65, 95% CI [−1.87, 0.70]). At the mean level of attitude, the effect was positive but non-significant (Effect = 0.61, Boot SE = 0.46, 95% CI [−0.29, 1.51]). At high levels of attitude, however, source expertise had a significant positive effect on knowledge (Effect = 1.80, Boot SE = 0.63, 95% CI [0.54, 3.05]).

These conditional effects (see Table 4; Fig. 3) indicate that source expertise increased objective knowledge only among participants with highly favorable attitudes toward gluten-free diets. Thus, H2 was supported.

Moderated moderation analysis

To test H3, we conducted a moderated moderation analysis using PROCESS Model 3 with objective knowledge as the dependent variable, source expertise as the independent variable, pre-existing attitude as the first moderator, and knowledge state (uncertain vs. misinformed) as the second moderator.

The overall model explained 14.1% of the variance in objective knowledge ($R^2 = 0.141$, $F(9, 93) = 1.70$, $p = .100$). The three-way interaction among source expertise, pre-existing attitude, and knowledge state was not statistically significant ($B = 0.33$, SE = 0.91, $t = 0.36$, $p = .722$; 95% CI [−1.48, 2.13]). Thus, H3 was not supported.

Exploratory subgroup analyses

Although H3 was not supported, we conducted exploratory subgroup analyses to examine whether the interaction between source expertise and pre-existing attitude differed across knowledge states (uncertain vs. misinformed), as indicated by the knowledge state indicator.

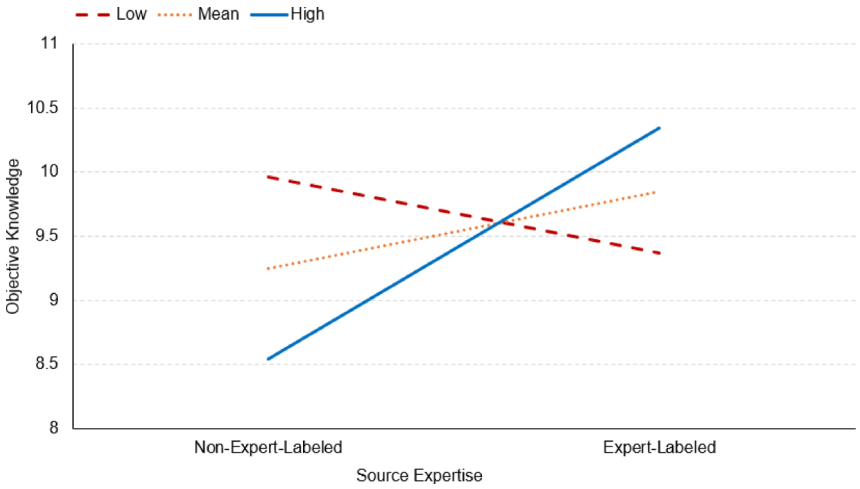


Fig. 3. Objective knowledge as a function of source expertise and pre-existing attitude. Lines represent low (− 1 SD), mean, and high (+ 1 SD) levels of pre-existing attitude.

Pre-existing attitude	Effect	Boot SE	Boot LLCI	Boot ULCI
Low (−1SD = 1.66)	−0.59	0.65	−1.87	0.70
Mean (M = 2.77)	0.61	0.45	−0.29	1.50
High (+ 1SD = 3.88)	1.80	0.63	0.54	3.05

Table 4. Conditional effects of source expertise on objective knowledge at levels of pre-existing attitude. Note. Bootstrap standard errors and 95% confidence intervals are based on 5,000 samples.

		DV: Objective knowledge				
		B	SE	*t*	*p*	R^2
Uncertain	Evaluation of information search	0.32	0.27	1.20	0.235	0.18
	Trust in ChatGPT	0.63	0.34	1.86	0.067	
	Source expertise	−2.80	1.54	−1.82	0.074	
	Pre-existing attitude	−0.89	0.45	−1.98	0.052	
	Source × Attitude	1.20	0.55	2.16	0.034*	
Misinformed	Evaluation of information search	0.26	0.45	0.59	0.563	0.10
	Trust in ChatGPT	−0.19	0.52	−0.37	0.711	
	Source expertise	−1.04	2.75	−0.38	0.708	
	Pre-existing attitude	−0.24	0.52	−0.46	0.650	
	Source × Attitude	0.67	0.79	0.85	0.405	

Table 5. Exploratory subgroup analyses: Source Expertise × Pre-existing Attitude predicting Objective Knowledge.

We tested the model separately for uncertain and misinformed participants (Table 5). Among uncertain participants ($n = 72$), the interaction between source expertise and pre-existing attitude was statistically significant ($B = 1.20$, SE = 0.55, $t = 2.16$, $p = .034$). The model explained 18.4% of the variance in objective knowledge ($R^2 = 0.184$, $F(5, 66) = 2.97$, $p = .018$). In contrast, among misinformed participants ($n = 31$), the interaction was not significant ($B = 0.67$, SE = 0.79, $t = 0.85$, $p = .405$), and the model as a whole did not significantly predict objective knowledge ($R^2 = 0.103$, $F(5, 25) = 0.57$, $p = .719$).

These subgroup findings descriptively align with the hypothesized pattern, showing that the interaction between source expertise and pre-existing attitude emerged among uncertain participants but not among misinformed participants. However, these results should be considered exploratory rather than confirmatory.

Discussion

This study examined how source expertise, pre-existing attitudes, and knowledge states jointly shape objective knowledge outcomes when individuals verify health information using generative AI. By integrating dual-process models of persuasion with theories of epistemic trust and distributed cognition, it advances understanding of

how individuals evaluate and learn from AI-supported information environments. The design simulated a real-world verification process by combining exposure to mixed-accuracy social media content with an AI-assisted search, capturing how people encounter, check, and update health information in digital contexts. In such settings, users do not merely retrieve information but engage in epistemic monitoring, relying on AI as a supportive informational partner while remaining vigilant toward its reliability[3,4].

Conditional effects of expertise

Source expertise alone did not significantly increase objective knowledge. Even after an AI-assisted search, participants encountering expert-labeled posts did not show higher levels of objective knowledge than participants exposed to non-expert-labeled posts. This result aligns with prior evidence that source credibility cues primarily shape evaluative responses such as perceived truth or trustworthiness rather than objective knowledge outcomes[26]. In other words, expertise cues may guide how information is judged but not necessarily how it is reflected in objective knowledge. This pattern suggests that in information environments where users verify mixed content using AI-assisted search, expertise appears to guide judgment but does not reliably translate into higher post-verification objective knowledge. Expertise cues may therefore function primarily as heuristic signals that orient attention rather than consistently supporting objectively assessed knowledge.

A significant interaction between source expertise and pre-existing attitudes clarified this pattern. Expert-labeled posts were associated with higher levels of factual knowledge only among participants with favorable attitudes toward gluten-free diets. This selective effect indicates that expertise is not universally persuasive but functions as a conditional catalyst activated when users are already motivated to engage more deeply. Favorable attitudes provide the motivational resources for systematic processing, consistent with dual-process models emphasizing that elaboration depends on both ability and motivation[8]. Related research in misinformation correction also finds that expert sources are most effective when recipients' beliefs align with the corrective message[16] and fact-checking was shaped by both credibility perceptions and prior belief alignment[13]. Meta-analytic evidence confirms that the persuasive impact of source credibility systematically depends on recipients' motivational orientations[43]. Taken together, these results suggest that expertise serves as a processing gatekeeper rather than a direct enhancer of learning—triggering epistemic effort only when motivational conditions favor engagement. Related evidence suggests that expert feedback may alter the weighting of supportive versus disconfirming information during belief updating, particularly in health contexts involving contested claims[44]. This perspective reinforces the view that expertise shapes how information is integrated, rather than directly enhancing factual accuracy.

Knowledge states: uncertainty versus misinformation

Although the predicted three-way interaction among expertise, attitudes, and knowledge state was not statistically significant, exploratory analyses revealed a consistent conceptual pattern. Uncertain individuals, starting without strong commitments, were more receptive to expertise cues; misinformed individuals, anchored in inaccurate beliefs, showed little change even after AI-assisted verification. This contrast aligns with evidence that misinformation resists correction due to motivated reasoning[19,20], whereas uncertainty allows cognitive flexibility[21,22]. More recent work similarly finds that corrective messages are particularly effective among those with uncertain or weakly held prior beliefs[45]. Together with the conceptual framework proposed by Schulz and Nakamoto[17,18], these exploratory patterns are consistent with prior accounts linking uncertainty to epistemic openness and misinformation to motivated resistance. Understanding these boundaries helps clarify why some individuals integrate expertise cues into their knowledge while others dismiss them. Importantly, knowledge state is treated here as a contextual boundary condition rather than a primary explanatory mechanism, and the present findings should not be interpreted as evidence for stable epistemic traits or categorical differences between individuals. Accordingly, conclusions regarding H3, are tentative and intended to inform future, more highly powered research.

Accordingly, the primary contribution of the study lies in the demonstrated interaction between source expertise and pre-existing attitudes, while analyses involving knowledge state serve to contextualize and qualify these effects. The latter are intended to inform theory development and future research rather than to provide definitive tests.

Extending persuasion models to AI-mediated communication

This study extends classical dual-process models such as the ELM[8] and HSM[9] to health communication contexts in which AI-assisted verification is available. These models, typically developed in interpersonal and media-based persuasion[46,47], assume that source expertise influences persuasion primarily through its interaction with motivation. The current findings suggest that these mechanisms also apply when users engage in AI–assisted information search. Recent work suggests that digital and AI environments alter how credibility cues are processed[48,49]. Our results support this view, showing that human–AI information processing involves both traditional motivational factors and new forms of epistemic vigilance toward the AI's reliability. Expertise effects thus remain conditional on users' motivational and epistemic states, even when generative AI provides corrective opportunities. Overall, this study extends dual-process theories into human–AI information processing by showing that expertise effects depend on motivational and epistemic conditions in contexts where AI provides opportunities for verification.

Practical implications and design insights

From an applied perspective, the findings underscore the limits of relying on expertise labels alone in mixed-accuracy environments. These cues were associated with higher levels of objective knowledge only among individuals already favorable to the topic, suggesting that communicators should combine source attribution

with interactive features that promote elaboration – such as evidence visualization, consensus framing, or prompts encouraging users to justify their reasoning. From a practical perspective, our results indicate that measures that are designed to highlight the expertise or credibility of a source of information could be less effective, or ineffective[50] when the target audience already holds strong opposing beliefs.

Second, the differences between uncertain and misinformed users highlight the value of adaptive AI communication. Generative AI systems could be designed to diagnose users' epistemic states and adjust responses accordingly: offering scaffolding and evidence synthesis for uncertain users, and confidence calibration or corrective dialogue for the misinformed. Moreover, emphasizing transparency and domain-specific credentials may foster appropriate trust calibration and prevent overreliance on AI outputs[5,41].

Limitations and future directions

Several limitations warrant attention. First, ChatGPT was not manipulated as an experimental factor but served as a shared verification affordance, the present study does not allow causal claims about the effects of AI use per se, but rather about how established credibility and motivational processes operate when AI-assisted verification is available. Within this design, the AI–assisted search task may have introduced uncontrolled variance in participants' motivation or cognitive load. Standardized prompts or comparison with non-AI search environments could help clarify these effects in future research.

Second, the distinction between uncertainty and misinformation was conceptually meaningful but not statistically evident, likely reflecting sample size constraints. Given the modest number of participants ($N = 103$) and the complexity of the moderated moderation model, the study likely had limited statistical power to detect small three-way interaction effects. Accordingly, the non-significant three-way interaction should be interpreted cautiously. In addition, analytic approaches that more directly model changes in objective knowledge may help clarify these boundary conditions when applied in larger samples. Future studies should test this boundary condition with larger, longitudinal or process-tracing designs that capture how individuals move between epistemic states as they search, verify, and update information.

Third, while pre-existing attitudes and knowledge states are conceptually distinct, some empirical overlap was observed, partly reflecting the use of an indicator based on response patterns and difference scores. Evaluative commitment may increase belief confidence regardless of factual accuracy, suggesting the need for refined measurement approaches that disentangle motivational and epistemic components of information processing. Relatedly, the operationalization of knowledge state reflects response patterns within a specific task context rather than enduring epistemic dispositions. While this approach aligns with recent conceptual distinctions between uncertainty and misinformation, future research should employ more direct and independent measures of epistemic confidence, belief strength, and response tendencies.

Fourth, source expertise was operationalized through an expert label rather than verified medical authority. Participants' perceived expertise of the source was not directly assessed, which may limit the extent to which the findings generalize to confirmed domain expertise.

Finally, the sample was drawn from a single cultural context, which limits the generalizability of the findings. Cross-cultural replications would provide valuable insight into whether the observed mechanisms hold across different cultural and technological settings. In addition, although participants' evaluations of the search experience and their trust in ChatGPT were included as covariates, other factors such as digital literacy, search motivation, and perceived task difficulty may also shape objective knowledge outcomes. Future research should include these variables to better specify the conditions under which AI-assisted knowledge acquisition occurs.

Conclusion

In sum, this study demonstrates that expertise effects in AI-mediated health communication are conditional rather than universal. Users' motivation and epistemic state determine whether expertise cues translate into objective knowledge. These dynamics illustrate a form of distributed epistemic processing, where credibility, effort, and objective knowledge emerge from collaboration between humans and AI rather than from message features alone. By bridging persuasion theory, cognitive psychology, and AI design, this research contributes to a more comprehensive understanding of how generative AI can support informed—rather than merely influenced—health decision-making[51–54].

Data availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Appendix 1: Items of the Objective Knowledge Regarding Gluten-Free Diet

1.Gluten is a protein found in wheat as well as in grains like barley, oats, and various other foods (T).
2.Gluten is associated with diseases such as diabetes and brain disorders, even in people without celiac disease or gluten sensitivity (F).
3.Gluten can cause severe bloating and fatigue in gluten-sensitive individuals (T).
4.Symptoms of gluten sensitivity persist even after a gluten-free diet (F).
5.Gluten-free diets are essential for people with diagnosed celiac disease (T).
6.A gluten-free diet requires strict avoidance of all gluten-containing foods and vigilance against cross-contamination (T).
7.Simply avoiding gluten can lead to weight loss (F).
8.'Wheat-free' is the same as gluten-free (F).

9.People with celiac disease are advised to follow a gluten-free diet to avoid complications such as osteoporosis, iron deficiency anemia, diarrhea, and bloating (T).
10.Natural and artificial flavors can contain gluten (T).
11.Gluten can be completely removed from grains like wheat and barley through certain processing techniques, making them safe for consumption by individuals with celiac disease (F).
12.Fermented products like soy sauce and sourdough bread are naturally gluten-free because the fermentation process breaks down the gluten proteins (F).
13.Certain over-the-counter medications and supplements may contain gluten as an inactive ingredient, which must be avoided by individuals on a strict gluten-free diet (T).
14.A gluten-free diet can potentially reduce the risk of cardiovascular diseases in the general population, even among those without gluten intolerance (F).
15.A gluten-free diet can be easily achieved by simply avoiding all processed foods (F).

Received: 3 December 2025; Accepted: 5 March 2026
Published online: 12 March 2026

References

1. Sun, X. et al. Trusting the search: unraveling human trust in health information from Google and ChatGPT. *arXiv:2403.09987*. (2024).
2. Yun, H. S. & Bickmore, T. Online health information–seeking in the era of large language models: Cross-sectional web-based survey study. *J. Med. Internet. Res.* **27**, e68560 (2025).
3. Risko, E. F. & Gilbert, S. J. Cognitive offloading. *Trends Cogn. Sci.* **20**(9), 676–688 (2016).
4. Sperber, D. et al. Epistemic vigilance. *Mind Lang* **25**(4), 359–393 (2010).
5. Cabrera, Á. A., Perer, A. & Hong, J. I. Improving human-AI collaboration with descriptions of AI behavior. *Proc. ACM Hum.-Comput. Interact.* **7**(CSCW1), 1–21 (2023).
6. Spitzer, P. et al. Human delegation behavior in human-AI collaboration: The effect of contextual information. *Proc. ACM Hum.-Comput. Interact.* **9**(2), 1–28 (2025).
7. Lerner, A., O'Bryan, T. & Matthias, T. Navigating the gluten-free boom: The dark side of gluten free diet. *Front. Pead.* **7**, 414 (2019).
8. Petty, R. E. & Cacioppo, J. T. The elaboration likelihood model of persuasion. In Advances in experimental social psychology (Vol. 19, 123–205). Academic. (1986).
9. Chaiken, S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *J. Pers. Soc. Psychol.* **39**(5), 752 (1980).
10. Pornpitakpan, C. The persuasiveness of source credibility: A critical review of five decades' evidence. *J. Appl. Soc. Psychol.* **34** (2), 243–281 (2004).
11. Susmann, M. W. & Wegener, D. T. The independent effects of source expertise and trustworthiness on retraction believability: The moderating role of vested interest. *Mem. Cognit.* **51**(4), 845–861 (2023).
12. Tormala, Z. L. & Petty, R. E. Source credibility and attitude certainty: A metacognitive analysis of resistance to persuasion. *J. Consum. Psychol.* **14**(4), 427–442 (2004).
13. Liu, X., Qi, L., Wang, L. & Metzger, M. J. Checking the fact-checkers: The role of source type, perceived credibility, and individual differences in fact-checking effectiveness. *Communication Res.* **52** (6), 719–746 (2025).
14. Schweiger, S. & Cress, U. Attitude confidence and source credibility in information foraging with social tags. *PLoS One* **14**(1), e0210423 (2019).
15. Van Strien, J. L., Kammerer, Y., Brand-Gruwel, S. & Boshuizen, H. P. How attitude strength biases information processing and evaluation on the web. *Comput. Hum. Behav.* **60**, 245–252 (2016).
16. Wang, Y. Debunking misinformation about genetically modified food safety on social media: Can heuristic cues mitigate biased assimilation?. *Sci. Commun.* **43**(4), 460–485 (2021).
17. Schulz, P. J. & Nakamoto, K. *Worse Than Ignorance: The Challenge of Health Misinformation* (Cambridge University Press, 2024).
18. Schulz, P. J. & Nakamoto, K. Understanding health knowledge failures: Uncertainty versus misinformation. *Sci. Rep.* **15**(1), 23867 (2025).
19. Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D. & Rich, R. F. Misinformation and the currency of democratic citizenship. *J. Polit.* **62**(3), 790–816 (2000).
20. Nyhan, B. & Reifler, J. When corrections fail: The persistence of political misperceptions. *Polit. Behav.* **32** (2), 303–330 (2010).
21. Chinn, C. A. & Brewer, W. F. The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Rev. Educ. Res.* **63**(1), 1–49 (1993).
22. Li, J. & Wagner, M. W. The value of not knowing: Partisan cue-taking and belief updating of the uninformed, the ambiguous, and the misinformed. *J. Commun.* **70**(5), 646–669 (2020).
23. Metzger, M. J. & Flanagin, A. J. Credibility and trust of information in online environments: The use of cognitive heuristics. *J. Pragmat.* **59**, 210–220 (2013).
24. Nadarevic, L., Reber, R., Helmecke, A. J. & Köse, D. Perceived truth of statements and simulated social media postings: An experimental investigation of source credibility, repeated exposure, and presentation format. *Cogn. Res. Princ. Implic.* **5**(1), 1–16 (2020).
25. Wertgen, A. G. & Richter, T. Source credibility modulates the validation of implausible information. *Mem. Cognit.* **48**(8), 1359–1375 (2020).
26. Mang, V., Fennis, B. M. & Epstude, K. Source credibility effects in misinformation research: A review and primer. *advances. in/psychology* https://doi.org/10.56296/aip00028 (2024).
27. Vraga, E. K. & Bode, L. Using expert sources to correct health misinformation in social media. *Sci. Commun.* **39**(5), 621–645 (2017).
28. Eagly, A. H. & Chaiken, S. *The psychology of attitudes* (Harcourt brace Jovanovich college, 1993).
29. Haugtvedt, C. P. & Petty, R. E. Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes. *J. Pers. Soc. Psychol.* **63**(2), 308 (1992).
30. Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N. & Cook, J. Misinformation and its correction: Continued influence and successful debiasing. *Psychol Sci Public Interest* **13**(3), 106–131 (2012).
31. Walter, N. & Tukachinsky, R. A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it?. *Commun. Res.* **47**(2), 155–177 (2020).
32. Ajzen, I. The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* **50**(2), 179–211 (1991).
33. Maio, G. R., Verplanken, B. & Haddock, G. The psychology of attitudes and attitude change. (2018).
34. El Khoury, D., Balfour-Ducharme, S. & Joye, I. J. A review on the gluten-free diet: Technological and nutritional challenges. *Nutrients* **10** (10), 1410 (2018).

35. Vraga, E. K. & Bode, L. Correction as a solution for health misinformation on social media. *Am. J. Public Health*. **110** (S3), S278–S280 (2020).
36. Chua, A. Y. & Banerjee, S. Intentions to trust and share online health rumors: An experiment with medical professionals. *Comput. Hum. Behav.* **87**, 1–9 (2018).
37. Sainsbury, K. & Mullan, B. Measuring beliefs about gluten free diet adherence in adult coeliac disease using the theory of planned behaviour. *Appetite* **56** (2), 476–483 (2011).
38. Leffler, D. A. et al. Factors that influence adherence to a gluten-free diet in adults with celiac disease. *Dig. Dis. Sci.* **53**(6), 1573–1581 (2008).
39. Sachdeva, R. A Scale to Assess the Efficacy of Consumer Decision Making. *IUP J. Mark. Manage.* **14**(2), 8-25 (2015).
40. Jian, J.-Y., Bisantz, A. M. & Drury, C. G. Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* **4**, 53–71. https://doi.org/10.1207/S15327566IJCE0401_04 (2000).
41. McGrath, M. J., Lack, O., Tisch, J. & Duenser, A. Measuring trust in artificial intelligence: Validation of an established scale and its short form. *Front. Artif. Intell.* **8**, 1582880 (2025).
42. Hayes, A. F. Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Commun. Monogr.* **85**(1), 4–40 (2018).
43. Kumkale, G. T., Albarracín, D. & Seignourel, P. J. The effects of source credibility in the presence or absence of prior attitudes: Implications for the design of persuasive communication campaigns. *J. Appl. Soc. Psychol.* **40**(6), 1325–1356 (2010).
44. Garcia-Arch, J., Barberia, I., Rodríguez-Ferreiro, J. & Fuentemilla, L. Authority brings responsibility: Feedback from experts promotes an overweighting of health-related pseudoscientific beliefs. *Int. J. Environ. Res. Public Health.* **19**(22), 15154 (2022).
45. Dan, V. & Coleman, R. "I'll change my beliefs when I see it": Video fact checks outperform text fact checks in correcting misperceptions among those holding false or uncertain pre-existing beliefs. *Commun. Res.* **52**(6), 778–802 (2025).
46. Briñol, P. & Petty, R. E. Elaboration and validation processes: Implications for media attitude change. *Media Psychol.* **18**(3), 267–291 (2015).
47. O'Keefe, D. J. The elaboration likelihood model. *The SAGE handbook of persuasion: Developments in theory and practice*, 137–149. (2013).
48. Lu, L., Tormala, Z. L. & Duhachek, A. How AI sources can increase openness to opposing views. *Sci. Rep.* **15**(1), 17170 (2025).
49. Matz, S. C. et al. The potential of generative AI for personalized persuasion at scale. *Sci. Rep.* **14**(1), 4692 (2024).
50. Oeldorf-Hirsch, A., Schmierbach, M., Appelman, A. & Boyle, M. P. The ineffectiveness of fact-checking labels on news memes and articles. *Mass. Communication Soc.* **23** (5), 682–704 (2020).
51. Grinschgl, S. & Neubauer, A. C. Supporting cognition with modern technology: Distributed cognition today and in an AI-enhanced future. *Front. Artif. Intell.* **5**, 908261 (2022).
52. Krosnick, J. A. & Petty, R. E. Attitude strength: An overview. *Attitude strength*, 1–24. (2014).
53. Petty, R. E., Cacioppo, J. T. & Schumann, D. Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *J. Consum. Res.* **10**(2), 135–146 (1983).
54. Swire-Thompson, B., DeGutis, J. & Lazer, D. Searching for the backfire effect: Measurement and design considerations. *J. Appl. Res. Mem. Cognit.* **9**(3), 286–299 (2020).

Author contributions

Conceptualization: J.O., P.J.S.Methodology: J.O., P.J.S.Investigation: C.M., J.K.Data curation: J.O.Formal analysis: J.O.Visualization: J.O.Writing – original draft: J.O.Writing – review and editing: J.O., P.J.S.Project administration: P.J.S.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to P.J.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026