Information Processing and Management 60 (2023) 103203

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Comparing button-based chatbots with webpages for presenting fact-checking results: A case study of health information

Xianglin Zhao, Li Chen, Yucheng Jin *, Xinzhi Zhang

Hong Kong Baptist University, Hong Kong, China

ARTICLE INFO

Keywords:
Health misinformation
Chatbot
User interface
Fact-checking
Conversational agent
User study

ABSTRACT

The unprecedented proliferation of online health misinformation poses a potential threat to public health. In recent times, several fact-checking organizations have adopted chatbots to present fact-checking results. However, it is unclear whether chatbots are more appropriate than traditional fact-checking websites for presenting these rigorous corrective messages. To answer this question, we compared a button-based chatbot with a traditional webpage for *presenting* fact-checking results. As fact-checkers' expertise cues could influence users' perception of fact-checking, we also considered the effect of expertise cues in our study. We conducted a 2 (interaction type: webpage vs. chatbot) × 2 (expertise cue: non-highlighted vs. highlighted) between-subjects online experiment ($N = 308$). The results show that the chatbot leads to higher perceived ease of use, which in turn increases the effectiveness of fact-checking. The highlighted expertise cue tends to decrease users' intention to use, especially when they interact with the webpage. Finally, we discuss the feasibility of using chatbots to disseminate fact-checking content and several design implications for the creation of an effective tool to fact-check health information.

1. Introduction

Health misinformation is defined as *"information contrary to the scientific community's epistemic consensus regarding a phenomenon"* (Swire-Thompson & Lazer, 2019). Widespread health misinformation can confuse the audience, lead to misperceptions about health issues, and cause risk-taking behaviors, resulting in detrimental effects on health (Nan et al., 2021). For example, some articles on health claim that drinking alcohol can prevent and cure COVID-19. Approximately 500 Iranians believed these articles and died after consuming toxic alcohol during the COVID-19 pandemic (Delirrad & Mohammadi, 2020).

Fact-checking is a significant and feasible solution to combat misinformation (Bode & Vraga, 2018; Hameleers & Van der Meer, 2020). To reduce potential health threats, it is essential to obtain insights on channels for proper and timely communication of verified health information to the public. During the COVID-19 pandemic, widespread rumors posed new challenges to the dissemination of health facts, prompting public healthcare departments and fact-checking organizations to launch chatbots for debunking rumors and responding to public concerns in a timely manner. For example, the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) launched chatbots to assess infection risk, share official information and facts, and offer emotional support (Amiri & Karahanna, 2022; Miner et al., 2020). In comparison with traditional webpages, chatbots provide a more interactive way to fact-check health information (Skjuve & Brandtzæ g, 2018), which could increase user

* Corresponding author.
E-mail addresses: csxlzhao@comp.hkbu.edu.hk (X. Zhao), lichen@comp.hkbu.edu.hk (L. Chen), yuchengjin@hkbu.edu.hk (Y. Jin), xzzhang2@hkbu.edu.hk (X. Zhang).

https://doi.org/10.1016/j.ipm.2022.103203
Received 15 June 2022; Received in revised form 18 November 2022; Accepted 22 November 2022
Available online 13 December 2022

satisfaction, accessibility, and ease of use (Amiri & Karahanna, 2022). This proactive interaction has also been proven to have a positive effect in correcting misconceptions, such as reducing vaccine hesitancy (Altay et al., 2021).

Chatbots have shown considerable promise for use in healthcare services. An increasing amount of evidence has demonstrated the active role of chatbots in facilitating disease prevention (Amato et al., 2017), chronic disease management (Roca et al., 2020), and mental health support (Vaidyam et al., 2019). However, the adoption of chatbots to combat health misinformation is still in its infancy. In comparison with traditional fact-checking webpages that display a large amount of information on a single page, using chatbots to obtain information requires more interaction and individuals may experience higher barriers (Han & Lee, 2022). Therefore, as an emerging tool for fact-checking, chatbots continue to face many challenges with respect to user acceptability, functionality, and usability (Almalki & Azeez, 2020).

In the context of health information, source expertise plays an important role in the perception of information quality and information acceptance (Hovland et al., 1953; Petty et al., 1981). For an efficient appraisal of source expertise, people tend to use simple inspections based on salient cues. For example, individuals are more likely to accept the health-related advice of a communicator whose title is "Dr." rather than "Mr." (Crisci & Kassinove, 1973). Studies have explored the manipulation of the visual cues of chatbots (e.g., avatars, identities) to signal source expertise; such cues are also referred to as expertise cues (Go & Sundar, 2019; Liew et al., 2021). Studies have demonstrated that these manipulations have positive effects on users' perceived competence of chatbots (Dai & MacDorman, 2018), persuasion (Ischen et al., 2020; Parmar et al., 2018), and users' trust toward chatbots (Parmar et al., 2018; Sah et al., 2011) in certain contexts. However, it remains unknown whether highlighting expertise cues is helpful for users when they obtain fact-checking results for health information, especially in conversations with chatbots. Therefore, we investigate the following three research questions:

RQ1: How does the *interaction type* (webpage vs. chatbot) influence user acceptance of the fact-checking tool and correction effectiveness?

RQ2: How does the *expertise cue* (non-highlighted vs. highlighted) influence user acceptance of the fact-checking tool and correction effectiveness?

RQ3: Do the effects of *interaction type* on fact-checking depend on the *expertise cue*?

In this regard, we conducted a 2 (*interaction type*: webpage vs. chatbot) × 2 (*expertise cue*: non-highlighted vs. highlighted) between-subjects study to evaluate the fact-checking user interface by using the Technology Acceptance Model (TAM) (Venkatesh et al., 2003); we evaluated perceived ease of use, perceived usefulness, correction effectiveness, intention to check, and intention to use. We collected 308 valid responses and used Structural Equation Modeling (SEM) to analyze the collected data. The results show that the button-based chatbot leads to higher perceived ease of use and increases correction effectiveness when debunking false health claims. Furthermore, the highlighted expertise cue could reduce participants' intention to check health information and use the fact-checking tool, especially in the case of the webpage.

To the best of our knowledge, this study is the first to empirically compare chatbots with webpages for *presenting* fact-checking information on health-related topics. This study makes several conceptual contributions to the user experience design of a fact-checking tool. We believe that this empirical study can also provide practical insights for designing an effective chatbot to combat health misinformation.

2. Related work

In this section, we briefly review studies on topics related to our study: health misinformation and fact-checking, fact-checking tools, and the expertise cue in fact-checking.

2.1. Health misinformation and fact-checking

Health misinformation generally consists of one or more health-related claims that are inconsistent with scientific consensus, potentially harming public health and misleading people in their health-related decisions (Nan et al., 2021). As a specific type of misinformation, it can be created through oversimplification, misrepresentation, over-dramatization in media coverage (Lewandowsky et al., 2012; Thomas et al., 2017), or deliberate use of pseudoscientific or poor quality evidence, misleading claims, and personal anecdotes (Al Khaja et al., 2018; Teoh, 2019). Individuals without a professional background may not be able to discern unreliable health-related messages because most of these messages are not ostensibly false but often subtly misleading (Nan et al., 2021). Therefore, laypeople may rely on professionals to obtain reliable health information.

As a primary strategy to combat misinformation, fact-checking is the process of evaluating verifiable claims made in public statements (Brandtzaeg & Følstad, 2017), the main form of which is *"presenting a corrective message that establishes that the prior message was misinformation"* (Chan et al., 2017). Several meta-analyses have indicated that fact-checking or debunking messages can effectively mitigate the threat posed by health misinformation to public health (Carey et al., 2022; Chan et al., 2017; Porter & Wood, 2021; Walter et al., 2021). Since the launch of the nonprofit FactCheck.org[1] in 2013, a growing number of fact-checking organizations brought together a wide variety of professionals to provide informative fact-checking or debunking content in order to help the public combat misinformation (Graves & Mantzarlis, 2020). The Duke Reporters' Lab stated that the number of fact-checking outlets had reached 188 in 2019, four times the number in 2014 (Stencel, 2019). Some well-known fact-checking websites,

[1] https://www.factcheck.org

such as Snopes[2] and AFP Fact Check,[3] have been playing an important role in checking health misinformation. For example, since the beginning of the COVID-19 pandemic, Snopes has received numerous queries about hydroxychloroquine and other rumors on "cures" (Leskin, 2020).

2.2. Fact-checking tools

Fact-checking websites serve as important tools with established editorial credentials to communicate facts to the online community by proactively investigating the veracity of claims and refuting dubious ones (Pal & Loke, 2019). Interventions for addressing misinformation can take many forms (Zhang et al., 2021), such as warning labels (Zhang et al., 2021), simple rebuttals (van der Meer & Jin, 2020), and factual elaboration (van der Meer & Jin, 2020). However, warning labels are not as effective as debunking or corrective messages in mitigating the effects of misinformation (Oeldorf-Hirsch et al., 2020; Walter & Murphy, 2018). This is because such labels do not provide the reasons for claims being assessed as false or alternative factual information, and may inadvertently backfire by improving familiarity through repetition (Peter & Koch, 2016; Schwarz et al., 2016). Furthermore, as Chan et al. (2017) revealed, "a detailed debunking message positively correlated with the misinformation-persistence effect". Debunking content can provide more recallable detail, which can reduce false beliefs in a more sustainable manner (Ecker et al., 2020). Therefore, the consumption of detailed fact-checking results provided by fact-checking organizations remains a fairly well-established and effective method to help the public correct false beliefs.

Several studies have investigated several design factors that could influence the effectiveness of fact-checking on these websites, such as presentation format (e.g., simple "false-tag" retractions vs. detailed refutations), content type (e.g., video vs. text), and message tone (humorous vs. non-humorous) (Ecker et al., 2020; Young et al., 2018). However, few studies have investigated the effects of interaction type (e.g., conversational interaction) on the communication of fact-checking results. Marcondes et al. (2019) conducted a survey to explore the application of chatbots in combating misinformation and found that the application of chatbots for this purpose was not common. Since the onset of the global COVID-19 pandemic, an increasing number of healthcare institutions and fact-checking organizations have discovered the potential value of chatbots in communicating health information because chatbots can simultaneously serve a large number of users and provide quick responses and necessary guidance.

In recent years, chatbots have become a popular way of providing health information and services (Athota et al., 2020; Brixey et al., 2017; Oh et al., 2017). Since the start of the COVID-19 pandemic, numerous fact-checking organizations have scrambled to employ chatbots to provide the public with fact-checking results and reliable information on the pandemic (Almalki & Azeez, 2020; Amiri & Karahanna, 2022). The WHO and the International Fact-Checking Network launched the WHO Health Alert chatbot (Walwema, 2021) and the FactChat chatbot (Leskin, 2020) via WhatsApp and gained a large number of users in a short period.

A study showed that a chatbot helped a healthcare department combat misinformation on COVID-19 vaccinations, resulting in a significant increase in participants' vaccination intentions (Altay et al., 2021). Nevertheless, employing chatbots to provide fact-checking results remains in its infancy. Amiri and Karahanna (2022) collected 13 use cases of chatbots that had been launched for fact-checking during the COVID-19 pandemic and found that text-based chatbots that interact via predetermined choice and response options (button-based) had become the predominant type of fact-checking chatbots. There are two main reasons for the widespread adoption of button-based chatbots. First, button-based chatbots can be developed at a lower cost and can be deployed rapidly, thus meeting the needs of most fact-checking organizations that use chatbots as an alternative channel for disseminating information. Second, button-based chatbots are more reliable than natural-language-based chatbots when communicating structured messages (Maniou & Veglis, 2020; Veglis & Maniou, 2019). Most studies of fact-checking chatbots have focused on system design and usability (Altay et al., 2021; Amiri & Karahanna, 2022; Roque et al., 2021; Siedlikowski et al., 2021). Few studies have investigated user acceptance of chatbots or the fact-checking effectiveness of chatbots in comparison with traditional webpages.

2.3. The expertise cue in fact-checking

The information source is considered to be responsible for communication and accountable for the provided information (Nowak & Biocca, 2003), thereby triggering heuristics and enabling individuals to effortlessly assess the credibility of information, especially in situations involving difficult judgment (Sundar, 2008). The expertise cue is often considered the main element determining source credibility (Liao & Fu, 2014). A typical mental shortcut is that information provided by sources with high expertise is considered more credible than information from sources with low expertise (Go et al., 2014). When information is provided by authorities or domain experts, signals that they are professionals can increase receivers' attention, processing effort, and positive perception, potentially leading to greater attitude change or behavioral compliance (Hovland & Weiss, 1951; Petty & Cacioppo, 1984). For example, people are more likely to follow the advice of psychologists when psychologists use the title "Dr." rather than the title "Mr." (Crisci & Kassinove, 1973). Vraga and Bode (2017) noted that in the correction of health misinformation on social media, expert sources are perceived to be more credible than laypeople sources.

When the interaction type is changed from a webpage to a chatbot, audiences may perceive the chatbot to be a fully responsive entity and assume that the fact-checking results originate from the chatbot rather than the professional fact-checkers behind it (Ischen

[2] https://www.snopes.com

[3] https://factcheck.afp.com

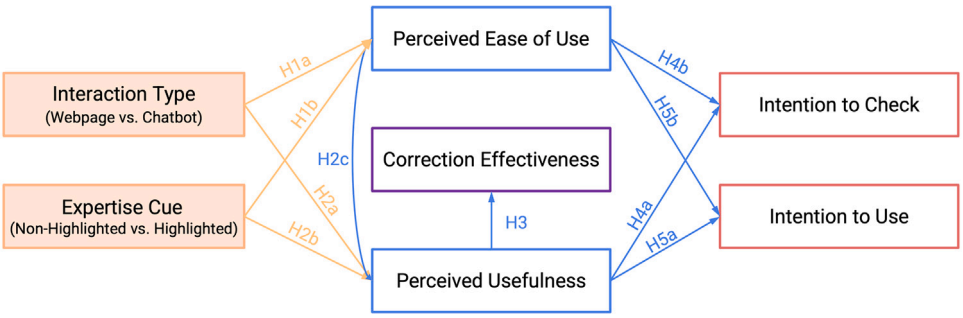

Fig. 1. Hypothetical research model illustrating the proposed research hypotheses.

et al., 2020). According to the social response theory (also called the "computers are social actors paradigm") (Nass & Moon, 2000), individuals may anthropomorphize conversational agents or chatbots, treating them as social actors (Ho et al., 2018; Nass & Moon, 2000). More specifically, a fact-checking chatbot may be considered an intelligent fact-checker rather than a mediator in the communication of verified health information. Thus, a change in the interaction type may affect audiences' judgment of verified information sources, which will have an unknown effect on information credibility and correction effectiveness.

According to Graves and Cherubini (2016), fact-checkers mainly consist of reporters, reformers, and experts. As the debunking of health misinformation relies on the highly specialized judgment of experts with domain knowledge, the role of fact-checkers in this context is usually performed by experts. Fact-checking websites typically specify fact-checkers' titles and backgrounds as evidence on their webpages to reflect their rigor and robustness and to increase information credibility (Pal & Loke, 2019). The fact-checkers' titles can be seen as an expertise cue. In conversational systems, agents' figures (avatars), names or identity labels, and descriptors in conversation are also considered expertise cues, which could influence the perceived source expertise of chatbots (Liew et al., 2021). Studies on chatbots for product recommendations and healthcare have revealed the positive effect of these expertise cues on purchase intention and platform trust (Go & Sundar, 2019; Liew et al., 2021; Niß en et al., 2022). Inspired by these studies, we aimed to manipulate avatars and professional titles as the expertise cue to explore the effect of this factor on fact-checking effectiveness for two interaction types: webpage and chatbot.

3. Research hypotheses

Given the context of this study, the interaction type and the expertise cue could influence user acceptance of fact-checking and correction effectiveness. We evaluated the design of the fact-checking tool based on the Technology Acceptance Model (TAM) (Davis, 1989). Fig. 1 shows our hypothetical research model. In this subsection, we review studies of the constructs measured in our model and propose corresponding research hypotheses.

Perceived Ease of Use. Perceived ease of use is defined as *"the extent to which users believe that using a system is free from cognitive effort"* (Saadé & Bahli, 2005). In this study, the interaction type may affect audiences' perceived ease of use in accessing fact-checking results, thereby influencing their behavioral intention. The reading of fact-checking results as an information processing activity requires the audience to expend their limited cognitive resources (Stanovich, 1990). The operation of the chatbot may require additional interactions, which may increase the cognitive effort in the reading process. For example, Nguyen et al. (2022) found that the chatbot interface resulted in higher cognitive effort than the menu-based interface in two information-searching tasks. Another study compared a chatbot for frequently asked questions (FAQs) with a webpage for FAQs and found that chatbot users experienced higher barriers (Han & Lee, 2022). However, in comparison with reading and processing the entire information in detail to determine whether to accept a message, audiences invest a relatively small amount of cognitive effort to make a decision based on heuristic cues (e.g., expertise cues) (Go et al., 2014). When the interface highlights the expertise cue, the audience may receive a signal that the information is credible, potentially avoiding the effort required for verification (Ratneshwar & Chaiken, 1991). Therefore, we propose the following hypotheses:

H1a: The chatbot will result in lower perceived ease of use than the webpage.

H1b: The highlighted expertise cue will result in higher perceived ease of use than the non-highlighted expertise cue.

Perceived Usefulness. Perceived usefulness is generally defined as *"the individual's belief that using a particular system would enhance the job performance"* (Davis, 1989). The TAM posits that perceived ease of use positively influences perceived usefulness of a system (Davis, 1989). In the case of fact-checking tools, perceived usefulness reflects the extent to which the content provided by the tool can be useful for correcting false beliefs.

Reading fact-checking results can be viewed as a process of persuasion. In communication research, persuasion is the process of judging whether a message is useful and convincing (Dillard et al., 2007). Zarouali et al. (2021) compared websites with chatbots for the delivery of news containing conflicting views and found that individuals agreed to a greater extent with a counter-attitudinal news story delivered by the chatbot. Furthermore, Ischen et al. (2020) found that the chatbot resulted in higher levels of persuasion than a website in the context of product recommendation.

Based on these findings, we hypothesize that the audience perceives fact-checking results from the chatbot to be more persuasive, which in turn results in higher perceived usefulness. In addition, the audience may trust the words of professionals to a greater extent, and many studies have confirmed that individuals rely on the expertise cue as a shortcut to assess the information credibility (Chaiken, 1980; Crisci & Kassinove, 1973; Vraga & Bode, 2017). Therefore, we hypothesize that:

H2a: The chatbot will result in higher perceived usefulness than the webpage.

H2b: The highlighted expertise cue will result in higher perceived usefulness than the non-highlighted expertise cue.

H2c: Perceived ease of use will have a positive effect on perceived usefulness.

Correction Effectiveness.

In our study, correction effectiveness refers to the percentage of false claims that the audience can correct after reading the fact-checking results. We consider the changes in beliefs or attitudes (e.g., the judgment of the claims) as a valid measure of correction effectiveness, as suggested by Swire et al. (2017) and Fennis and Stroebe (2015). Several studies have shown that correction effectiveness can be predicted based on perceived usefulness (Dillard et al., 2007; Noar et al., 2020, 2010). Therefore, we propose the following hypothesis:

H3: Perceived usefulness will have a positive effect on correction effectiveness.

Behavioral Intentions. Behavioral intention is defined as *"individuals' motivation or willingness to exert effort to perform the target behavior"* (Holden & Karsh, 2010). In this study, we analyze users' behavioral intentions in terms of both intention to check and intention to use. Intention to check refers to the audience's willingness to actively verify information in the future after reading the fact-checking results. Intention to use reflects a user's desire to use a system in the future (Teo & Zhou, 2014). According to the TAM, perceived usefulness and perceived ease of use are the two most significant determinants in explaining behavioral intentions (Davis, 1989). Lehto and Oinas-Kukkonen (2015) noted that perceived usefulness of the information positively influences the user's intention to use an interface for obtaining information and continue using it in the future. Therefore, we propose the following hypotheses:

H4a: Perceived usefulness will positively influence users' intention to check.

H4b: Perceived ease of use will positively influence users' intention to check.

H5a: Perceived usefulness will positively influence users' intention to use.

H5b: Perceived ease of use will positively influence users' intention to use.

4. Method

To investigate the differences between the webpage and the chatbot in the presentation of fact-checking results, we imagined a fact-checking organization called "Hawkeye" and developed a webpage and a button-based chatbot to present verified information on behalf of the organization. Furthermore, we manipulated the expertise cue for the fact-checking results by showing the fact-checkers' avatars and their titles in bold type to investigate the effect of the expertise cue on correction effectiveness and user acceptance. The details of the experiment design are described in the following subsections.

4.1. Design manipulations

4.1.1. Manipulation of interaction type

In this experiment, we implemented two types of interaction (*webpage* and *chatbot*) for presenting the fact-checking results. We adopted the general design of a fact-checking website (Pal & Loke, 2019) to create our fact-checking webpage. Specifically, the page first shows the veracity indicator (Fig. 2(A)) stating the level of truthfulness (e.g., "false", "mostly false", "mostly true", "true"), and then shows the claim summary (Fig. 2(B)) as brief explanations or justifications. Furthermore, the detailed fact-checking results (Fig. 2(C)) explain the process used for investigating the claims and provide more detailed explanations and justifications for veracity. The fact-checkers' titles and avatars are shown as the expertise cue for the fact-checking results (Fig. 2(D)). These structural headings, visual elements, and layout are commonly found on fact-checking websites, enabling participants to experience a fact-checking webpage in a more practical setting.

Similar to the structure of the webpage, the chatbot first shows the veracity indicator (Fig. 3(A)), and then the claim summary (Fig. 3(B)), followed by questions about false claims (e.g., "Can staying up late cause pneumothorax?") in the form of a list of buttons (Fig. 3(C)). Both the webpage and the chatbot show the same fact-checking content. The chatbot buttons are created based on claim and fact pairs presented on the webpages. The first button, "Which false claims were found in this news story?", provides users with an overview of all the facts, and is followed by a list of buttons that help users navigate to the details of each verified claim.

The interaction design for this chatbot follows the Q&A style and does not support typing by a user. We chose this design for several reasons. First, in practice, many fact-checking organizations continue to implement a chatbot with predetermined response options and simple tasks as an alternative method for accessing existing information. This method meets the need for quick deployment and does not require sophisticated programming (Amiri & Karahanna, 2022). Second, this study focuses on comparing different interaction types for presenting fact-checking results, which does not require a more flexible input such as typing input. Third, it is not appropriate to perform a comparison between a chatbot supporting text input and a static webpage in terms of user input capability. Therefore, we finally adopted the simple button-based chatbot design to ensure that the experimental conditions are close to the real-world setting.

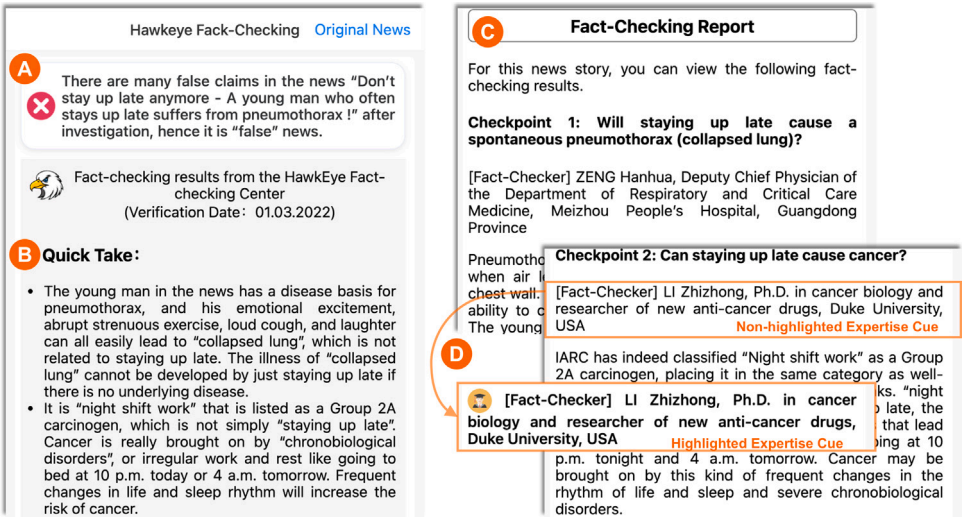

Fig. 2. Screenshots of the webpage showing (A) the veracity indicator, (B) snippets of the claim summary, (C) detailed fact-checking results, and (D) the expertise cue for fact-checking. The content shown in the screenshots was translated from Chinese to English for easy understanding.

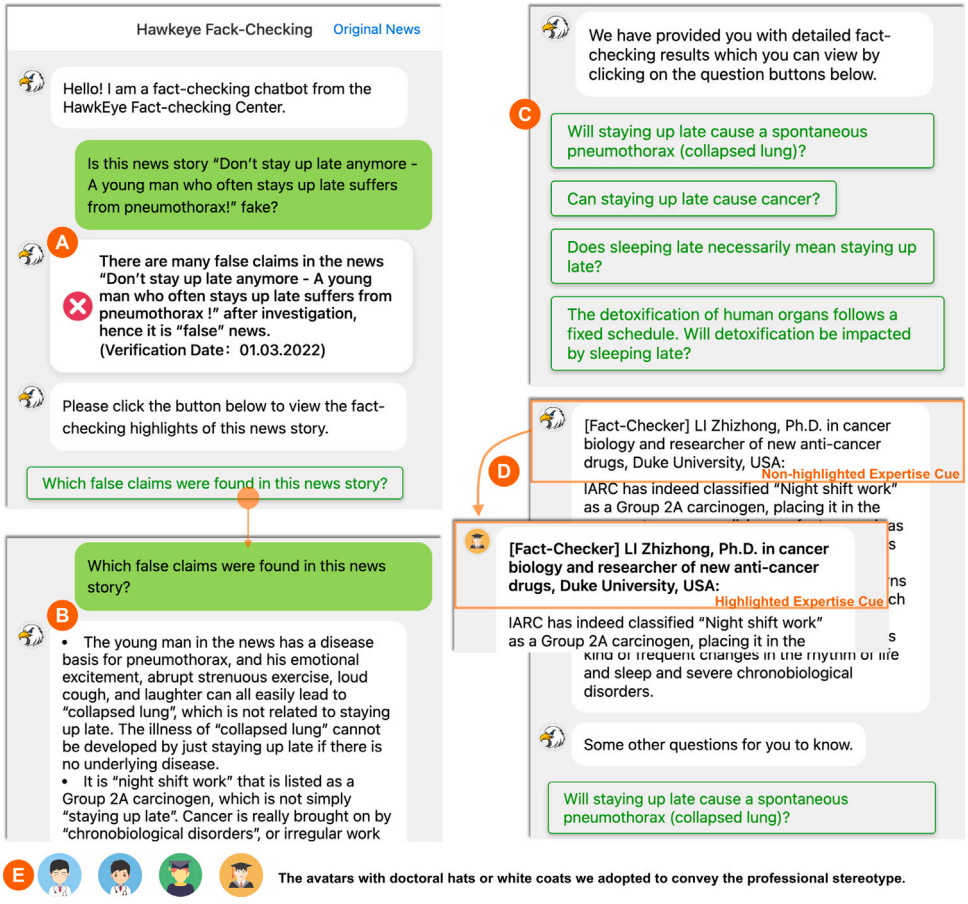

Fig. 3. Screenshots of the chatbot showing (A) the veracity indicator, (B) snippets of the claim summary, (C) the button group that presents the verified claims, (D) the expertise cue for fact-checking, and (E) the professional avatars. The content shown in the screenshot was translated from Chinese to English for easy understanding.

4.1.2. Manipulation of expertise cue

Changing the visual representations of expertise (e.g., avatars) is a primary way to manipulate the expertise cue (EC) (Go & Sundar, 2019; Liew et al., 2021; Niß en et al., 2022). For the scenario of the highlighted expertise cue, we adopted avatars (Fig. 3(E)) based on the fact-checkers' titles to convey the typical stereotypes for professionals. For example, we used an avatar with a white coat and a stethoscope to suggest that the fact-checker was a doctor. In general, the health information presented contains the identity and title of the fact-checker because health-related facts are typically provided by health professionals (Bautista et al., 2021). Thus, instead of showing or hiding this identity information, we chose to highlight the professional title of the fact-checker by using bold typeface. Bold typeface has been proven to be the most effective text-highlighting technique (Strobelt et al., 2015). For the scenario of the non-highlighted expertise cue, we showed the fact-checkers' titles in a regular font.

4.2. Stimuli and materials

The topic that we chose for health misinformation was the negative effects of "staying up late" on health; this topic attracted the attention of a greater number of participants than the other two topics in our pilot study: "artificial sweetener" and "weight management". A fact-checking article generally provides multi-faceted information to correct false claims about a topic. However, we could not find a single news article that contains multiple false claims about "staying up late". Therefore, we modified a news article by incorporating four common false claims about "staying up late" verified by a well-known Chinese fact-checking organization called JiaoZhen.[4]

The corrective messages for each false claim in the fact-checking report were also collected from the "JiaoZhen" platform. Four fact-checkers with backgrounds in professional healthcare provided detailed justifications and explanations for these claims. Based on the verified information, we created a fact-checking report by following a standard format, which provides a brief summary of facts at the beginning and then corrects each false claim based on a detailed justification and explanation. Following fact-checking website practices, we included each fact-checker's title in the report.

4.3. Procedure

To test our hypotheses in a real-world setting, we conducted a 2 (interaction type: webpage vs. chatbot) × 2 (expertise cue: non-highlighted vs. highlighted) between-subjects study online. We organized a lucky draw session with 30 e-vouchers as prizes for participants who had completed the entire experiment; each voucher was worth CNY 100 (≈ USD 15). The Research Ethics Committee of Hong Kong Baptist University approved our study on March 17, 2022. To test the procedure for this online experiment and the stimulus materials, we conducted a pilot study with 78 undergraduate students (gender: female 78.2%, male 21.8%; age: Mean = 19.0, SE = 0.09) before the formal study. Based on the results, we adjusted the number and wording of the questions in the questionnaire and modified the stimulus content and guidance information. The formal experiment was conducted at several universities in China from April 14, 2022 to April 25, 2022. The average time required for participants to complete all the activities in the experiment was 12 min. Fig. 4 illustrates our formal procedure for the experiment, which consisted of five main steps:

1. *Reading the consent form and the introduction.* We asked the participants to first sign a consent form before participating in the study and to then read a brief introduction before the experiment; the introduction included content such as the study goal, procedure, estimated duration, ethical statement, and compensation. Individuals were allowed to participate in the experiment through personal computers or mobile devices; 97.3% of the participants used their mobile devices, and 2.7% used their personal computers.
2. *Filling out the pre-study questionnaire.* The participants answered demographic questions and judged the truth of six test claims on the topic of "staying up late".
3. *Reading the news article containing false claims about "staying up late".* The participants read the news article on "staying up late" that we had created. To ensure that they spent an adequate amount of time reading the article, we asked them to read the article carefully before proceeding to the next step. To capture their natural reading behavior, we avoided setting a minimum dwell time on the news page for the participants.
4. *Reading the fact-checking results.* Based on a randomly assigned experimental condition, the participants read the fact-checking results on the webpage or the chatbot and with different expertise cues (non-highlighted or highlighted). As in the previous step, we did not set a minimum amount of time for the participants to read the fact-checking results.
5. *Filling out a post-task questionnaire.* The participants evaluated the fact-checking tool by filling out a post-task questionnaire, which included questions to measure aspects related to *perceived ease of use*, *perceived usefulness*, and *behavioral intentions* (i.e., intention to check and intention to use). Additionally, we asked the participants to judge the same six test claims that had been judged in the pre-task questionnaire. Following suggestions from previous studies (Ecker et al., 2015; Swire et al., 2017), we also asked the participants to complete a distraction task (finding 16 differences between two pictures in two minutes) before judging the six claims; this task could prevent the participants from mentally repeating the judgment they had made in the pre-task questionnaire.

[4] https://vp.fact.qq.com/

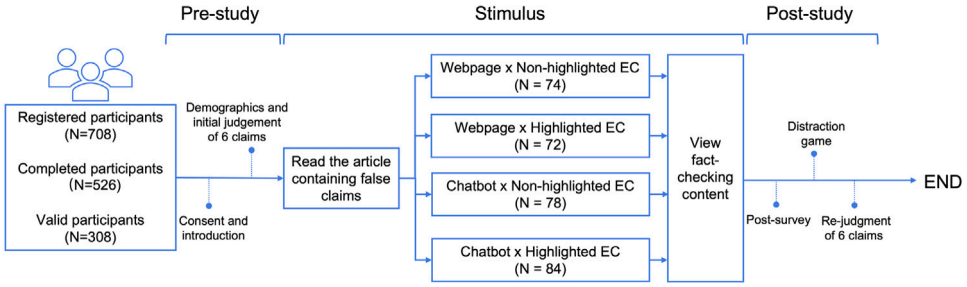

Fig. 4. Study procedure. Participants were randomly assigned to one of four conditions after they had read the article containing multiple false claims. EC is short for expertise cues.

Table 1
Demographic profile.

Property	Value	Sample	Percentage
Gender	Female	222	72.1%
	Male	85	27.6%
	Other	1	0.3%
Education	Bachelor's	254	82.5%
	Master's	38	12.3%
	PhD	6	1.9%
	Other	10	3.3%
Age	18–20	154	50.0%
	21–23	91	29.6%
	24–27	42	13.6%
	28 and above	21	6.8%

4.4. Participants

In all, 708 individuals registered for our formal experiment. We recruited these volunteers from several universities in China using social media and bulletin boards. During the experiment, 25.7% of the participants (N = 182) dropped out. Finally, 526 individuals participated in the entire experiment. To ensure good quality of the data, we filtered the user responses according to three criteria: (1) inattentive respondents (N = 196), i.e., those who failed to pass the attention check (e.g., "Please indicate option 5 (Strongly Agree) for this question" (DeSimone et al., 2015)) in the pre-task questionnaire or the post-task questionnaire; (2) extreme outliers with respect to the completion time (N = 3), who were detected based on the 3 × IQR (interquartile range) criterion; and (3) straightlining respondents (N = 19), i.e., those who provided identical responses to all the post-task questions. Finally, the responses of 308 participants were found to be valid and were used for analysis. Table 1 shows the demographic information for the participants whose responses were considered valid.

Participants were randomly assigned to one of the four conditions, and the sample size for the four groups was approximately the same: webpage with non-highlighted expertise cue, $N = 74$; webpage with highlighted expertise cue, $N = 72$; chatbot with non-highlighted expertise cue, $N = 78$; chatbot with highlighted expertise cue, $N = 84$. As the proportion of female participants (72.1%) in the sample was higher than that of male participants (27.6%), we further examined the gender differences between these four groups by using Fisher's exact test with Monte Carlo simulation. The result did not indicate significant differences ($p = .12$), indicating that the gender proportion was balanced for each condition.

4.5. Measures

We evaluated the fact-checking tool on the basis of several factors, such as perceived ease of use, perceived usefulness, correction effectiveness, users' intention to check health information, and users' intention to use the fact-checking tool. We assessed each factor by using questions validated by previous studies (see details in Table 3). We asked the participants to report their level of agreement with each question on a 5-point Likert scale. As all the participants were native Chinese speakers, we translated all the questions into simplified Chinese. To ensure good quality of the translations, we used the parallel translation method (Harkness et al., 2004): first, two authors of this paper independently translated the questionnaire, and then, they referred the items that could not be agreed on to the third author for the final decision. All the authors involved in translating the questions are native Chinese speakers and have professional-level proficiency in English.

To establish the validity of the constructs measured in our study, we performed a Confirmatory Factor Analysis (CFA) to meet the following cut-off values for different validity indexes: 0.7 for Cronbach's alpha,[5] 0.5 for average variance extracted (AVE) (Hair et al.,

[5] The cut-offs for Cronbach's alpha: 0.5 (poor), 0.6 (questionable), 0.7 (acceptable), 0.8 (good), 0.9 (excellent) (George & Mallery, 2019).

Table 2
Questions for measurement of correction effectiveness.

Claim	True/False
1. Staying up late can cause pneumothorax (collapsed lung).	False
2. Night shift work is a group 2A carcinogen (probable carcinogenic).	True
3. Statistically speaking, "disrupting the 24-hour body clock" is associated with a higher risk of cancer than drinking alcohol.	False
4. Circadian rhythm disruption can lead to drug resistance and reduced life span for cancer patients.	True
5. It is not an unhealthy sleep habit to go to bed at 2 a.m. and get up at 10 a.m. every day.	True
6. Human organs have a fixed detoxification schedule; staying up late can affect the detoxification time of body organs.	False

2010), and 0.5 for factor loading.[6] We excluded the question items that had low loadings ($R^2 < 0.5$) until the Average Variance Extracted (AVE) was greater than 0.5.[7] Consequently, we established the convergent and discriminant validity of the measured constructs. Convergent validity indicates whether the question items measure the same construct, and discriminant validity indicates whether two different constructs measure two different factors.

4.5.1. Perceived ease of use

To assess perceived ease of use, we used four items from a previous study (Venkatesh et al., 2003). Based on the CFA results, we excluded two items (Item 1 and Item 4) corresponding to perceived ease of use (Table 3) due to a poor factor loading value (< 0.5). Therefore, we considered perceived ease of use to be a single-item factor determined by Item 3 because a stable construct should have at least three items (Hair et al., 2010).

4.5.2. Perceived usefulness

In this study, perceived usefulness of the fact-checking tool differs from the perceived improvement in job performance after using a system as defined in the original TAM (Davis, 1989). In the context of fact-checking, perceived usefulness is an indicator of the extent to which the content provided by the tool is useful for fact-checking. Therefore, we used six questions developed by Dillard et al. (2007) to measure perceived usefulness. We asked the participants to evaluate the content provided by the fact-checking tool on the basis of six word pairs: "Not convincing–Convincing", "Not Believable–Believable", "Not sensible–Sensible", "Foolish–Wise", "Wrong–Right", and "Unimportant–Important". Based on the CFA results, Item 4, Item 5, and Item 6 (see Table 3) were removed due to a low factor loading value (< 0.5).

4.5.3. Correction effectiveness

Studies have shown that the corrective effect of fact-checking can be gauged from changes in beliefs or attitudes (Fennis & Stroebe, 2015; Swire et al., 2017). Therefore, we measured correction effectiveness by comparing the number of claims correctly judged before and after using our provided fact-checking tool. Specifically, we asked participants to judge the same six claims (Table 2) in the pre-task and post-task questionnaires. They judged each claim by choosing one of the following options: "True", "False", and "I don't know". The selection of "I don't know" was counted as an incorrect judgment of a statement's veracity. To balance the number of true and false claims, we ensured that three claims were faulty (myths) and the other three were correct (facts).

4.5.4. Intention to check

We used three questions from a previous study (Brodsky et al., 2021) to measure intention to check. This construct focuses on an individual's intention to actively verify information after viewing fact-checking results with the provided tool. We excluded Item 2 corresponding to intention to check (Table 3) due to a low factor loading value (< 0.5). As a construct cannot keep three items, we considered the intention to check to be a single-item factor determined by Item 1.

4.5.5. Intention to use

We measured intention to use based on the three validated questions (Venkatesh et al., 2003). This construct measures respondents' intention to use the fact-checking tool in the future using three question items.

5. Results

This section presents the results of Structural Equation Modeling (SEM) to simultaneously verify all the proposed research hypotheses about the effects of independent variables (IVs) on the dependent variables (DVs).

Based on the constructs validated by CFA, we performed SEM analysis by using the R package Lavaan 0.6-11.[8] The resulting model showed a good fit: $\chi^2(38) = 57.938$, $p < .05$; Root Mean Square Error of Approximation (RMSEA) = 0.041; 90% CI: [0.017, 0.062]; Comparative Fit Index (CFI) = 0.985, Tucker-Lewis Index (TLI) = 0.978.[9] Fig. 5 shows the model that represents the main

[6] The cut-offs for factor loading: 0.32 (poor), 0.45 (fair), 0.55 (good), 0.63 (very good), 0.71 (excellent) (Stevens, 2012).
[7] A suggested cut-off value for acceptable values for AVE is 0.5 (Brown, 2015).
[8] https://lavaan.ugent.be/
[9] Good values for the indices are: CFI > .96, TLI > .95, and RMSEA < .05 (Hu & Bentler, 1999)

Table 3
Post-task questionnaire.

Construct	Item	Factor loading
Perceived ease of use		
	1. My interaction with this fact-checking tool is clear and understandable.	
	2. It is easy for me to become skillful at using this fact-checking tool.	
	3. I found this fact-checking tool easy to use.	
	4. Learning to operate this fact-checking tool is not easy for me.	
Perceived usefulness (Cronbach's alpha = 0.890, AVE = 0.732)		
The fact-checking content provided by the fact-checking tool is:	1. not convincing/convincing.	0.517
	2. not believable/believable.	0.558
	3. not sensible/sensible.	0.590
	4. foolish/wise.	
	5. wrong/right.	
	6. unimportant/important.	
Intention to check		
	1. In the future, I will actively fact-check health-related news.	
	2. In the future, I will not try to verify the truthfulness of health-related news.	
	3. In the future, I will earnestly attempt to find out the truth about health-related news.	
Intention to use (Cronbach's alpha = 0.870, AVE = 0.692)		
	1. I intend to use this fact-checking tool in the future.	0.660
	2. I think I will use this fact-checking tool in the future.	0.747
	3. I plan to use this fact-checking tool in the future.	0.735

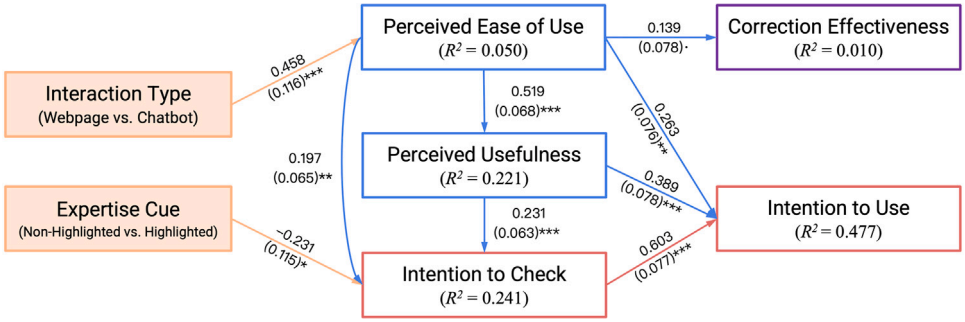

Fig. 5. Structural Equation Model (SEM) for the experiment. Significance levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $\cdot\ p < 0.10$. The numbers on the arrows refer to the β coefficient and the standard error of the causal relationship. R^2 is the proportion of variance explained by the model.

effects of the IVs (i.e., interaction type and expertise cue) on the DVs and the significant causal relationships between constructs. In addition, Fig. 6 shows the marginal effects of the two IVs on each DV, indicating how the values of a DV change with variation in the value of an IV (Norton et al., 2019).

5.1. Direct effects

Effect of Interaction Type. As shown in the model, the direct positive effect of the interaction type on perceived ease of use ($\beta = 0.458$, $p < .001$) suggests that using the chatbot was easier than using the webpage for reading fact-checking results. Based on the marginal effect on perceived ease of use (Fig. 6(a)), we found that it required less effort to consume fact-checking results through the chatbot than through the webpage, regardless of the expertise cue. Additionally, we did not find any other direct effects of the interaction type. Therefore, hypotheses **H1a** and **H2a** are not supported.

Effect of Expertise Cue. The expertise cue had a direct negative influence on intention to check ($\beta = -0.231$, $p < .05$), indicating that the participants were more reluctant to check the health information if the expertise cue for the fact-checking results was highlighted. The marginal effect on intention to check (Fig. 6(d)) suggests that the participants had considerably less intention to check the facts in the case of the highlighted expertise cue than with the non-highlighted expertise cue when using the webpage. The highlighted professional title and avatar of the fact-checkers appeared to have resulted in a decrease in participants' intention to check the facts. Moreover, as the expertise cue did not significantly influence perceived ease of use and perceived usefulness, hypotheses **H1b** and **H2b** are not supported.

Interaction Effects. To investigate the combined effect of the two IVs on the DVs, we also analyzed the interaction effect of the two IVs on all the DVs. Fig. 7 shows two significant interaction effects. Specifically, we found that the interaction effect on perceived usefulness was significant ($F(1, 304) = 4.60$, $p < .05$). As observed in Fig. 7(a), participants perceived higher usefulness with the chatbot than with the webpage when the expertise cue was highlighted. In comparison, the webpage resulted in higher perceived usefulness than the chatbot when the expertise cue was not highlighted. We also found a significant interaction effect on intention

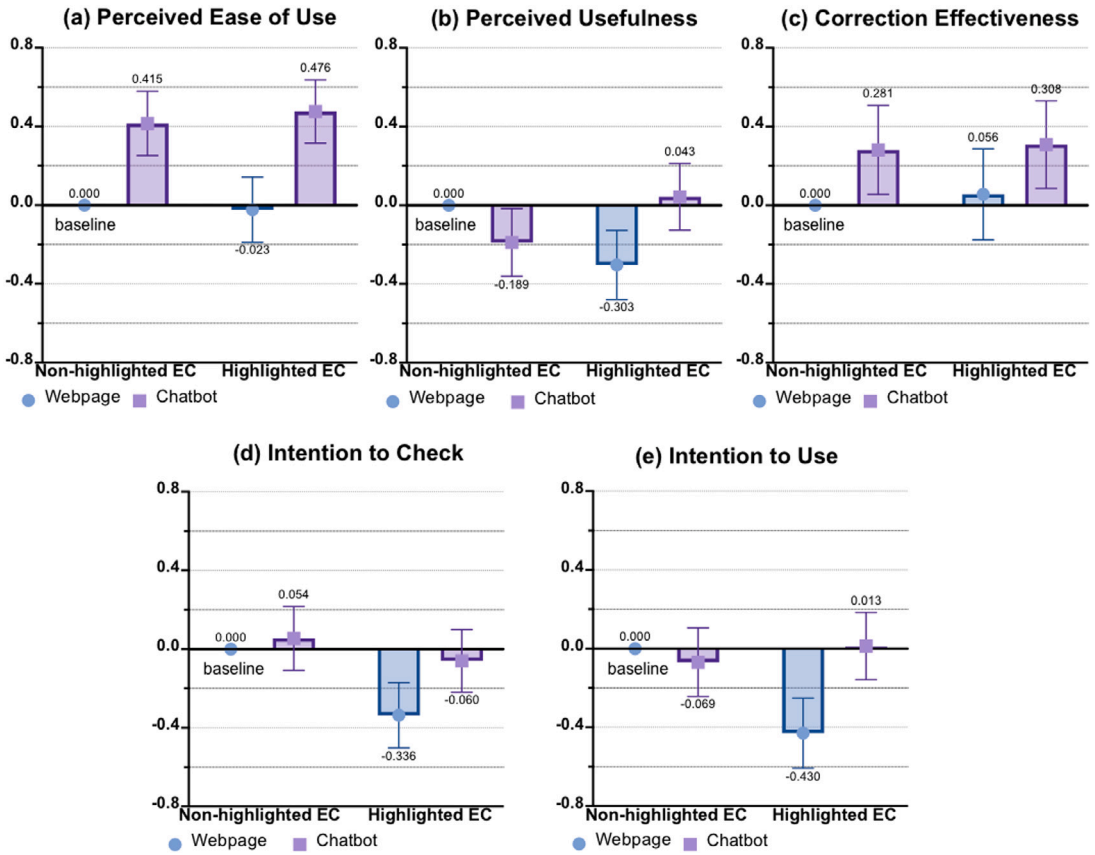

Fig. 6. Marginal effects of the interaction type and the expertise cue (EC) on five dependent variables (DVs). The effect of the "Webpage × Non-highlighted EC" is set to zero, and the numerical values represent the partial derivatives (slopes) of the regression equations with respect to each regressor. The error bars represent the standard errors.

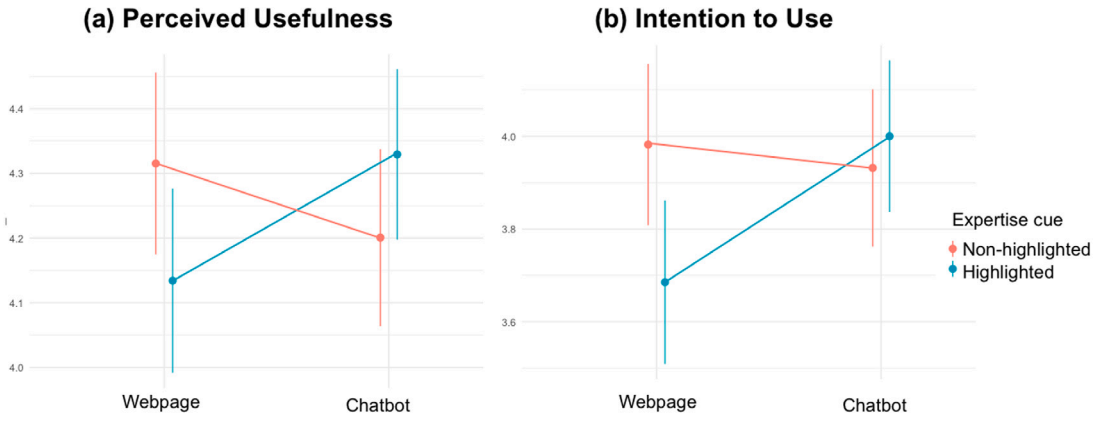

Fig. 7. Results of non-parametric factorial ANOVA. Two significant interaction effects exist between the interaction type and the expertise cue on (a) perceived usefulness and (b) intention to use. The error bars indicate the 95% confidence interval.

to use ($F(1304) = 7.16$, $p < .01$). The chatbot positively influenced intention to use if the expertise cue was highlighted. However, it appeared that the participants preferred the webpage over the chatbot for reading the fact-checking results in the case of the non-highlighted expertise cue. Therefore, the effects of interaction type on perceived usefulness and intention to use depend on the expertise cue.

Table 4
Descriptive analysis of all measured dependent variables.

Dependent variable	Webpage ×Non-Highlighted EC (N = 74) Mean (SD)	Webpage ×Highlighted EC (N = 72) Mean (SD)	Chatbot ×Non-Highlighted EC (N = 78) Mean (SD)	Chatbot ×Highlighted EC (N = 84) Mean (SD)
Perceived ease of use	3.88 (0.81)	3.86 (0.76)	4.19 (0.70)	**4.23 (0.77)**
Perceived usefulness	4.13 (0.62)	4.05 (0.70)	4.10 (0.72)	**4.29 (0.54)**
Correction effectiveness	1.51 (1.47)	1.57 (1.47)	1.79 (1.22)	**1.82 (1.44)**
Intention to check	4.05 (0.88)	3.75 (0.99)	**4.10 (0.81)**	4.00 (0.94)
Intention to use	3.98 (0.86)	3.69 (0.82)	3.93(0.68)	**4.00 (0.67)**

Note:
1. The highest value of each dependent variable is highlighted in bold.
2. The correction effectiveness value lies in the range between −6 and 6.

5.2. Indirect effects

The IVs also had several indirect effects on some DVs, such as correction effectiveness, perceived usefulness, intention to check, and intention to use.

Indirect Effect on Correction Effectiveness. The path [*interaction type → perceived ease of use → correction effectiveness*] indicates that the chatbot had an indirect positive effect on correction effectiveness ($\beta = 0.064$, $p = .101$). In other words, the chatbot increased the ease of fact-checking, which in turn improved the participants' performance in correcting false claims. The marginal effects on correction effectiveness (Fig. 6(c)) also show that the chatbot outperformed the webpage, regardless of the expertise cue. However, the amount of variance explained for correction effectiveness ($R^2 = 0.010$) is relatively small, which means that some influential factors of correction effectiveness were not investigated in our study. Therefore, hypothesis **H3** is not supported.

Indirect Effect on Perceived Usefulness. The model shows a path [*interaction type → perceived ease of use → perceived usefulness*] that indicates an indirect positive effect of interaction type on perceived usefulness. The combined total effect on perceived usefulness was 0.238 ($p < .001$), indicating that the use of the chatbot tended to increase perceived usefulness. However, based on the interaction effect on perceived usefulness, the positive effect of the chatbot depended on the highlighted expertise cue.

Indirect Effect on Intention to Check. The interaction type had several indirect effects as shown in the paths [*interaction type → (mediators) → intention to check*]. By calculating the total effects on intention to check, we found that the interaction type positively affected intention to check ($\beta = 0.145$, $p < .001$). Both paths have a common mediator, perceived ease of use, indicating that the participants who perceived fact-checking as less effortful were more likely to verify health information. Therefore, hypotheses **H4a** and **H4b** are supported.

Indirect Effect on Intention to Use. Several paths support the indirect effects of interaction type on intention to use (*interaction type → (mediators) → intention to use*]. The total effect on intention to use was 0.300 ($p < .001$). Perceived ease of use and perceived usefulness were the main mediators of the effect of interaction type on intention to use. In addition, the experience cue had an indirect negative effect on intention to use through intention to check ($\beta = -0.139$, $p < .05$). Therefore, hypotheses **H5a** and **H5b** are supported.

5.3. Summary of findings

To summarize our findings, we provide a descriptive analysis of the DVs based on the experimental conditions in the 2 × 2 between-subjects study in Table 4. In general, the chatbot × highlighted condition appears to have been the best design choice for presenting the fact-checking results. From the resulting model (Fig. 5), the chatbot positively influenced users' perceived ease of use with regard to the fact-checking tool, thereby increasing correction effectiveness. However, the highlighted expertise cue tended to result in a decrease in users' behavioral intentions, especially when the webpage was used.

6. Discussion

This study uses the TAM (Davis, 1989) to investigate the effects of two design variables of a fact-checking tool (i.e., interaction type and expertise cue) on various aspects of the fact-checking tool—perceived ease of use, perceived usefulness, correction effectiveness, intention to check, and intention to use. In this section, we discuss the major findings of this study that provide answers to the proposed research questions and present the design implications for presenting fact-checking results of health information.

6.1. RQ1: How does the interaction type influence user acceptance of the fact-checking tool and correction effectiveness?

In comparison with the webpage, the chatbot results in higher perceived ease of use. In other words, participants perceived the cognitive effort to be less when viewing the fact-checking results via the chatbot than via the webpage. According to the working memory theory (Baddeley & Logie, 1999; Miyake & Shah, 1997), humans have a limited capacity to store and process information within a set time. We argue that because a webpage often presents the full-text results at once, this may cause information overload and lead to an increase in receivers' cognitive effort. In comparison, a chatbot offers fact-checking results through several rounds

of dialogue, and each round shows one fact-checker's viewpoint. Chunking fact-checking results into different rounds of dialogue enables users to process information more easily. Studies have shown that the segmenting of information can reduce receivers' cognitive effort (Mayer & Chandler, 2001; Mayer & Moreno, 2003). Furthermore, the cognitive effort is closely related to the performance of reading and learning from texts (DeStefano & LeFevre, 2007), which could explain why the use of a chatbot results in higher correction effectiveness than the use of a webpage.

Although our results indicate that the chatbot results in higher perceived ease of use than the webpage, several studies have shown that the use of a chatbot resulted in higher cognitive load than the use of a webpage in information search tasks (Han & Lee, 2022; Nguyen et al., 2022). This discrepancy may be attributable to the various interaction capabilities enabled in chatbots and the different tasks performed. Unlike a chatbot that supports free-text input, the chatbot used in this study was restricted to button-based interaction, which could reduce the difficulty of interaction and prevent the chatbot from misunderstanding user input. Moreover, the task performed by the chatbot used in this study involved reading fact-checking results, which is easier than tasks related to information search.

Our model also shows how the interaction type can indirectly influence users' behavioral intentions. However, based on the marginal effects on users' behavioral intentions, the chatbot with the highlighted or non-highlighted expertise cue does not show significantly higher values than the baseline condition. Although the chatbot used in this study can reduce the cognitive effort required to view the fact-checking results, the intention to check health information also depends on other factors, such as users' initial judgment of the information (Edgerly et al., 2020) and media literacy (Lee & Ramazan, 2021). Therefore, we argue that using the chatbot alone may not be adequate to influence users' intentions to check health information and use a fact-checking tool.

6.2. RQ2: How does the expertise cue influence user acceptance of the fact-checking tool and correction effectiveness?

The results indicate that highlighting the expertise cue has a limited impact on user acceptance of the fact-checking tool and correction effectiveness, echoing the limited effects of source cues on fact-checking as reported by previous studies (Case et al., 2021; Wintersieck, 2017). Our results show that the highlighted expertise cue reduces users' intention to check, especially when they read the fact-checking results on the webpage. The negative effect of expertise cues on intention to check may be due to the backfire effect of expertise cues (Bohner et al., 2002). The backfire effect refers to the negative effects of highlighting an expertise cue on users' attitudes toward a message whose content does not meet users' expectations of the expertise cue. For example, in our study, some participants may not have been convinced by an expert's argument, and therefore, highlighting the expertise cue may have caused a negative reaction toward fact-checking health information.

6.3. RQ3: Do the effects of the interaction type on fact-checking depend on the expertise cue?

The interaction effects of the two IVs suggest that the expertise cue can influence the effect of interaction type on perceived usefulness and intention to use. We argue that the full-text results provided by a webpage enable users to quickly judge whether the fact-checking results are useful even when the expertise cue is not highlighted. However, a chatbot interface chunks the fact-checking results into multiple rounds of dialogue. Therefore, it is difficult for users to directly perceive the usefulness at the beginning of their interaction with the chatbot. Studies have shown that trust in chatbots can influence perceived usefulness of the messages they convey (Przegalinska et al., 2019; Rodríguez Cardona et al., 2021). Thus, participants are less likely to perceive the fact-checking results as useful if they do not trust the chatbot used. Nevertheless, a highlighted expertise cue could compensate for the lower perceived usefulness when using the chatbot. We speculate that the highlighted expertise cue helps users realize that these corrective messages are from the experts highlighted in the chatbot rather than from the chatbot itself, thereby reducing the effect of users' attitudes toward the chatbot on perceived usefulness of the fact-checking results.

A webpage with a highlighted expertise cue has an adverse effect on the intention to check and use (Figs. 6(d) and (e)). As the webpage presents full-text content, participants invest a greater amount of cognitive effort when processing text-heavy information. The highlighted expertise cue could increase the complexity of the webpage and attract participants' visual attention, which further increases cognitive overload and causes the user to expend more effort to look for the needed information (DeStefano & LeFevre, 2007).

Combining a chatbot with the highlighted expertise cue appears to be more effective. This combination achieves the highest value for all DVs except intention to check, which echoes the findings of a previous study reporting that the expertise cue fosters users' trust in an e-commerce chatbot, thereby increasing users' purchase intention (Liew et al., 2021).

6.4. Design implications

Overall, the chatbot results in higher perceived ease of use and correction effectiveness than the webpages. However, the advantages of the chatbot with regard to perceived usefulness and behavioral intentions are not evident and also depend on the expertise cue. Based on the results of our study, we present several practical implications for the design of fact-checking tools.

Design Implication #1: We suggest that a fact-checking webpage should avoid highlighting expertise cues and should limit visual complexity. We find that highlighting the expertise cue on a webpage leads to a decrease in perceived usefulness and behavioral intentions. Furthermore, the presentation of a large amount of information and many visual elements is often associated with a negative first impression of the content, resulting in a complete loss of the intention to read the content shown on the website (Crutzen et al., 2012). A higher visual complexity also negatively influences human cognition and emotion (Tuch et al.,

2009). To maintain a low visual complexity, designers can segment the fact-checking results and use the accordion user interface (UI) component to show and hide detailed content. The information overload caused by a text-heavy page inevitably increases the cognitive load for the user. This problem could be solved by highlighting the key points and refining the content.

Design Implication #2: We suggest that a fact-checking chatbot should highlight the expertise cue and avoid complex interactions. Unlike a webpage, which provides full-text content, a chatbot presents content in a step-by-step manner on receiving users' requests. Therefore, designers could employ the expertise cue to help individuals gain confidence about the credibility of the chatbot at the initial stage. From our study, possible ways to emphasize the expertise cue include showing professional avatars and highlighting the titles of experts. Other source cues that could influence user perceptions of news and chatbots, such as verbal style cues (Niß en et al., 2022), conversational cues (Go & Sundar, 2019), and bandwagon cues (Go et al., 2014), could also be used. Designers should also be aware that the complexity of interactions with chatbots can significantly increase the cognitive effort of users and reduce perceived ease of use (Han & Lee, 2022; Nguyen et al., 2022). Although free-text input provides more flexibility, it is not as simple as clicking a button and may introduce misunderstandings because of the chatbot's limited ability to understand natural language. Therefore, the use of predefined rules and button-based chatbots could ensure efficiency and reliability in delivering fact-checking results.

Design Implication #3: We suggest that the presentation of fact-checking results should be adapted according to the interaction type.

The presentation of fact-checking results with different interaction types can facilitate the accessibility of fact-checking services. However, designers must be cautious about directly including the corrective content from a webpage in a chatbot or directly adopting the design elements of a chatbot for a webpage. A mismatch between the presentation and the interaction may adversely affect fact-checking effectiveness and user experience. For example, chunking the content into several rounds of dialogue is more suitable than presenting the entire text in the chatbot.

6.5. Comparison with previous work

Studies of fact-checking chatbots have mainly focused on evaluating the interaction, functionality, and reliability of the chatbot (Maniou & Veglis, 2020; Roque et al., 2021). Differing from these studies, our study compared a chatbot with a traditional webpage for presenting fact-checking health information. The results of our study echo previous findings that chatbots could be promising tools for the dissemination of fact-checking results. For example, people are more likely to accept counter-attitudinal opinions delivered by a chatbot than by a webpage (Zarouali et al., 2021). Several studies have also shown that chatbots outperform traditional webpages in tasks beyond fact-checking, such as learning factual knowledge (Ruan et al., 2019), collecting data in surveys (Kim et al., 2019), and microtask crowdsourcing (Mavridis et al., 2019). The findings of our study also indicate that the chatbot results in higher correction effectiveness for users than the webpage.

Most of the studies investigating the effect of expertise cues have been conducted by considering a specific interaction type. For example, some studies of webpage use have found that an expertise cue can influence users' perception of information (Bhuiyan et al., 2021; Go et al., 2014; Naujoks & Benkenstein, 2020); and studies conducted with a chatbot have reported the effects of expertise cues on users' perceived competence (Dai & MacDorman, 2018), persuasion (Ischen et al., 2020; Parmar et al., 2018), and trust (Parmar et al., 2018; Sah et al., 2011). Unlike previous studies, our study investigated how different interaction types could influence the effect of expertise cues on user acceptance of fact-checking and correction effectiveness. Our findings reflect the need to adapt the expertise cue to the interaction type.

6.6. Limitations and future work

Our study has several limitations. *First*, we conducted the experiment by using only one modified article that contained four false claims about staying up late as stimulus material. Although this topic is one of the most popular health topics, as identified in our pilot study, our study cannot reveal whether health misinformation on other topics would produce different results. Future studies could use several health articles on different topics. *Second*, our use of the expertise cue in practical settings has some limitations with respect to real-world settings. This is because the results of fact-checking may include not only the justification and explanation from fact-checkers or experts but also the process and mechanism of verification. The description of these processes and mechanisms is important to ensure the neutrality and transparency of fact-checking organizations and to increase credibility. In future work, other visual cues should be explored to increase the credibility of the verified information delivered by a fact-checking chatbot. *Third*, the sample used in our study primarily consists of young adults from several universities, having a high level of education. However, the perception of misinformation may differ between younger adults and older adults due to differences in learning and memory (Marche et al., 2002). Studies of populations of different ages and educational backgrounds are necessary for future work. *Finally*, to ensure that all the participants involved in different interaction types viewed the same fact-checking results, we asked the chatbot users to click all the claim buttons. However, in practice, chatbot users may only check false claims that they are interested in.

7. Conclusion

The rise of button-based chatbots in fact-checking motivated us to investigate a case study that compares chatbots with webpages for presenting fact-checking results. We conducted an empirical user study to compare chatbots and webpages for presenting fact-checking results of health misinformation, with a non-highlighted or highlighted expertise cue. In general, our findings suggest that the chatbot plays a promising role in combating health misinformation because it results in better ease of use and correction effectiveness than a webpage. Although a button-based chatbot cannot trigger users' intention to check health information and use the fact-checking tool, it can guide users to explore fact-checking results according to their concerns and increase the effectiveness of fact-checking and perceived ease of use. Furthermore, the expertise cue has a direct negative effect on users' intention to check health information, thereby decreasing the intention to use the fact-checking tool. In addition, we find that the effects of the expertise cue on perceived usefulness and intention to use also depend on the interaction type. After discussing our findings, we provide several design implications for the presentation of fact-checking results with different interaction types and expertise cues. We believe that this work will make an important contribution to the interaction design of fact-checking tools.

CRediT authorship contribution statement

Xianglin Zhao: Methodology, Software, Formal analysis, Investigation, Writing – original draft. **Li Chen:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition. **Yucheng Jin:** Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Xinzhi Zhang:** Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported by the Hong Kong Baptist University IRCMS Project (IRCMS/19-20/D05).

References

Al Khaja, K. A., AlKhaja, A. K., & Sequeira, R. P. (2018). Drug information, misinformation, and disinformation on social media: A content analysis study. *Journal of Public Health Policy*, *39*(3), 343–357.

Almalki, M., & Azeez, F. (2020). Health chatbots for fighting COVID-19: A scoping review. *Acta Informatica Medica*, *28*(4), 241.

Altay, S., Hacquin, A.-S., Chevallier, C., & Mercier, H. (2021). Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions.. *Journal of Experimental Psychology: Applied*.

Amato, F., Marrone, S., Moscato, V., Piantadosi, G., Picariello, A., & Sansone, C. (2017). Chatbots meet eHealth: Automatizing healthcare. In *WAIAH@ AI* IA* (pp. 40–49).

Amiri, P., & Karahanna, E. (2022). Chatbot use cases in the Covid-19 public health response. *Journal of the American Medical Informatics Association*, *29*(5), 1000–1010.

Athota, L., Shukla, V. K., Pandey, N., & Rana, A. (2020). Chatbot for healthcare system using artificial intelligence. In *2020 8th international conference on reliability, infocom technologies and optimization (trends and future directions)* (pp. 619–622). IEEE.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In *Models of working memory: mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press, http://dx.doi.org/10.1017/CBO9781139174909.005.

Bautista, J. R., Zhang, Y., & Gwizdka, J. (2021). Healthcare professionals' acts of correcting health misinformation on social media. *International Journal of Medical Informatics*, *148*, Article 104375.

Bhuiyan, M. M., Whitley, H., Horning, M., Lee, S. W., & Mitra, T. (2021). Designing transparency cues in online news platforms to promote trust: Journalists' & consumers' perspectives. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–31.

Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, *33*(9), 1131–1140.

Bohner, G., Ruder, M., & Erb, H.-P. (2002). When expertise backfires: Contrast and assimilation effects in persuasion. *British Journal of Social Psychology*, *41*(4), 495–519.

Brandtzaeg, P. B., & Følstad, A. (2017). Trust and distrust in online fact-checking services. *Communications of the ACM*, *60*(9), 65–71.

Brixey, J., Hoegen, R., Lan, W., Rusow, J., Singla, K., Yin, X., Artstein, R., & Leuski, A. (2017). Shihbot: A facebook chatbot for sexual health information on HIV/AIDS. In *Proceedings of the 18th annual sigdial meeting on discourse and dialogue* (pp. 370–373).

Brodsky, J. E., Brooks, P. J., Scimeca, D., Galati, P., Todorova, R., & Caulfield, M. (2021). Associations between online instruction in lateral reading strategies and fact-checking COVID-19 news among college students. *AERA Open*, *7*, Article 23328584211038937.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Carey, J. M., Guess, A. M., Loewen, P. J., Merkley, E., Nyhan, B., Phillips, J. B., & Reifler, J. (2022). The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada. *Nature Human Behaviour*, *6*(2), 236–243.

Case, C., Eddy, C., Hemrajani, R., Howell, C., Lyons, D., Sung, Y.-H., & Connors, E. C. (2021). The effects of source cues and issue frames during COVID-19. *Journal of Experimental Political Science*, 1–10.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752.

Chan, M.-p. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, *28*(11), 1531–1546.

Crisci, R., & Kassinove, H. (1973). Effect of perceived expertise, strength of advice, and environmental setting on parental compliance. *The Journal of Social Psychology*, *89*(2), 245–250.

Crutzen, R., de Kruif, L., & de Vries, N. K. (2012). You never get a second chance to make a first impression: The effect of visual complexity on intention to use websites. *Interaction Studies*, *13*(3), 469–477.
Dai, Z., & MacDorman, K. F. (2018). The doctor's digital double: How warmth, competence, and animation promote adherence intention. *PeerJ Computer Science*, *4*, Article e168.
Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.
Delirrad, M., & Mohammadi, A. B. (2020). New methanol poisoning outbreaks in Iran following COVID-19 pandemic. *Alcohol and Alcoholism*.
DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*(2), 171–181.
DeStefano, D., & LeFevre, J.-A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, *23*(3), 1616–1641.
Dillard, J. P., Shen, L., & Vail, R. G. (2007). Does perceived message effectiveness cause persuasion or vice versa? 17 consistent answers. *Human Communication Research*, *33*(4), 467–488.
Ecker, U. K., Lewandowsky, S., Cheung, C. S., & Maybery, M. T. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language*, *85*, 101–115.
Ecker, U. K., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology*, *111*(1), 36–54.
Edgerly, S., Mourão, R. R., Thorson, E., & Tham, S. M. (2020). When do audiences verify? How perceptions about message and source influence audience verification of news headlines. *Journalism & Mass Communication Quarterly*, *97*(1), 52–71.
Fennis, B. M., & Stroebe, W. (2015). *The psychology of advertising*. Psychology Press.
George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge.
Go, E., Jung, E. H., & Wu, M. (2014). The effects of source cues on online news perception. *Computers in Human Behavior*, *38*, 358–367.
Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304–316.
Graves, L., & Cherubini, F. (2016). The rise of fact-checking sites in Europe. *Digital News Project Report*.
Graves, L., & Mantzarlis, A. (2020). Amid political spin and online misinformation, fact checking adapts. *The Political Quarterly*, *91*(3), 585–591.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis: A global perspective. Pearson Education.
Hameleers, M., & Van der Meer, T. G. (2020). Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers? *Communication Research*, *47*(2), 227–250.
Han, S., & Lee, M. K. (2022). FAQ chatbot and inclusive learning in massive open online courses. *Computers & Education*, *179*, Article 104395.
Harkness, J., Pennell, B.-E., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. *Methods for Testing and Evaluating Survey Questionnaires*, 453–473.
Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, *68*(4), 712–733.
Holden, R. J., & Karsh, B.-T. (2010). The technology acceptance model: Its past and its future in health care. *Journal of Biomedical Informatics*, *43*(1), 159–172.
Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). Communication and persuasion.
Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, *15*(4), 635–650.
Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.
Ischen, C., Araujo, T., van Noort, G., Voorveld, H., & Smit, E. (2020). "I am here to assist you today": The role of entity, interactivity and experiential perceptions in chatbot persuasion. *Journal of Broadcasting & Electronic Media*, *64*(4), 615–639.
Kim, S., Lee, J., & Gweon, G. (2019). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–12).
Lee, D. K. L., & Ramazan, O. (2021). Fact-checking of health information: The effect of media literacy, metacognition and health information exposure. *Journal of Health Communication*, *26*(7), 491–500.
Lehto, T., & Oinas-Kukkonen, H. (2015). Explaining and predicting perceived effectiveness and use continuance intention of a behaviour change support system for weight loss. *Behaviour & Information Technology*, *34*(2), 176–189.
Leskin, P. (2020). One of the internet's oldest fact-checking organizations is overwhelmed by coronavirus misinformation - and it could have deadly consequences.
Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131.
Liao, Q. V., & Fu, W.-T. (2014). Expert voices in echo chambers: Effects of source expertise indicators on exposure to diverse opinions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2745–2754).
Liew, T. W., Tan, S.-M., Tee, J., & Goh, G. G. G. (2021). The effects of designing conversational commerce chatbots with expertise cues. In *2021 14th international conference on human system interaction* (pp. 1–6). IEEE.
Maniou, T. A., & Veglis, A. (2020). Employing a chatbot for news dissemination during crisis: Design, implementation and evaluation. *Future Internet*, *12*(7), 109.
Marche, T. A., Jordan, J. J., & Owre, K. P. (2002). Younger adults can be more suggestible than older adults: The influence of learning differences on misinformation reporting. *Canadian Journal on Aging/la Revue Canadienne Du Vieillissement*, *21*(1), 85–93.
Marcondes, F. S., Almeida, J. J., & Novais, P. (2019). A short survey on chatbot technology: Failure in raising the state of the art. In *International symposium on distributed computing and artificial intelligence* (pp. 28–36). Springer.
Mavridis, P., Huang, O., Qiu, S., Gadiraju, U., & Bozzon, A. (2019). Chatterbox: Conversational interfaces for microtask crowdsourcing. In *Proceedings of the 27th ACM conference on user modeling, adaptation and personalization* (pp. 243–251).
Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*, *93*(2), 390.
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52.
van der Meer, T. G., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication*, *35*(5), 560–575.
Miner, A. S., Laranjo, L., & Kocaballi, A. B. (2020). Chatbots in the fight against the COVID-19 pandemic. *NPJ Digital Medicine*, *3*(1), 1–4.
Miyake, A., & Shah, P. (1997). Models of working memory. Colorado University at Boulder, Department of Psychology.
Nan, X., Wang, Y., & Thier, K. (2021). Health misinformation. In *The Routledge handbook of health communication* (pp. 318–332). Routledge.
Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103.
Naujoks, A., & Benkenstein, M. (2020). Expert cues: How expert reviewers are perceived online. *Journal of Service Theory and Practice*.
Nguyen, Q. N., Sidorova, A., & Torres, R. (2022). User interactions with chatbot interfaces vs. menu-based interfaces: An empirical study. *Computers in Human Behavior*, *128*, Article 107093.
Niß en, M., Rüegger, D., Stieger, M., Flückiger, C., Allemand, M., v Wangenheim, F., & Kowatsch, T. (2022). The effects of health care chatbot personas with different social roles on the client-chatbot bond and usage intentions: Development of a design codebook and web-based study. *Journal of Medical Internet Research*, *24*(4), Article e32630.
Noar, S. M., Barker, J., Bell, T., & Yzer, M. (2020). Does perceived message effectiveness predict the actual effectiveness of tobacco education messages? A systematic review and meta-analysis. *Health Communication*, *35*(2), 148–157.

Noar, S. M., Palmgreen, P., Zimmerman, R. S., Lustria, M. L. A., & Lu, H.-Y. (2010). Assessing the relationship between perceived message sensation value and perceived message effectiveness: Analysis of PSAs from an effective campaign. *Communication Studies*, *61*(1), 21–45.
Norton, E. C., Dowd, B. E., & Maciejewski, M. L. (2019). Marginal effects—quantifying the effect of changes in risk factors in logistic regression models. *JAMA*, *321*(13), 1304–1305.
Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *12*(5), 481–494.
Oeldorf-Hirsch, A., Schmierbach, M., Appelman, A., & Boyle, M. P. (2020). The ineffectiveness of fact-checking labels on news memes and articles. *Mass Communication and Society*, *23*(5), 682–704.
Oh, K.-J., Lee, D., Ko, B., & Choi, H.-J. (2017). A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. In *2017 18th IEEE international conference on mobile data management* (pp. 371–375). IEEE.
Pal, A., & Loke, C. (2019). Communicating fact to combat fake: Analysis of fact-checking websites. In *Proceedings of the 2019 international conference on information technology and computer communications* (pp. 66–73).
Parmar, D., Olafsson, S., Utami, D., & Bickmore, T. (2018). Looking the part: The effect of attire and setting on perceptions of a virtual health counselor. In *Proceedings of the 18th international conference on intelligent virtual agents* (pp. 301–306).
Peter, C., & Koch, T. (2016). When debunking scientific myths fails (and when it does not): The backfire effect in the context of journalistic coverage and immediate judgments as prevention strategy. *Science Communication*, *38*(1), 3–25.
Petty, R. E., & Cacioppo, J. T. (1984). Source factors and the elaboration likelihood model of persuasion. *ACR North American Advances*.
Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*(5), 847.
Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences*, *118*(37).
Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P., & Mazurek, G. (2019). In bot we trust: A new methodology of chatbot performance measures. *Business Horizons*, *62*(6), 785–797.
Ratneshwar, S., & Chaiken, S. (1991). Comprehension's role in persuasion: The case of its moderating effect on the persuasive impact of source cues. *Journal of Consumer Research*, *18*(1), 52–62.
Roca, S., Sancho, J., García, J., & Alesanco, Á. (2020). Microservice chatbot architecture for chronic patient support. *Journal of Biomedical Informatics*, *102*, Article 103305.
Rodríguez Cardona, D., Janssen, A., Guhr, N., Breitner, M. H., & Milde, J. (2021). A matter of trust? Examination of chatbot usage in insurance business. In *Proceedings of the 54th Hawaii international conference on system sciences* (p. 556).
Roque, G., Cavalcanti, A., Nascimento, J., Souza, R., & Queiroz, S. (2021). BotCovid: Development and evaluation of a chatbot to combat misinformation about COVID-19 in Brazil. In *2021 IEEE international conference on systems, man, and cybernetics* (pp. 2506–2511). IEEE.
Ruan, S., Jiang, L., Xu, J., Tham, B. J.-K., Qiu, Z., Zhu, Y., Murnane, E. L., Brunskill, E., & Landay, J. A. (2019). Quizbot: A dialogue-based adaptive learning system for factual knowledge. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13).
Saadé, R., & Bahli, B. (2005). The impact of cognitive absorption on perceived usefulness and perceived ease of use in on-line learning: An extension of the technology acceptance model. *Information & Management*, *42*(2), 317–327.
Sah, Y. J., Yoo, B., & Sundar, S. S. (2011). Are specialist robots better than generalist robots? In *Proceedings of the 6th international conference on human-robot interaction* (pp. 241–242).
Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, *2*(1), 85–95.
Siedlikowski, S., Noël, L.-P., Moynihan, S. A., & Robin, M. (2021). Chloe for COVID-19: Evolution of an intelligent conversational agent to address infodemic management needs during the COVID-19 pandemic. *Journal of Medical Internet Research*, *23*(9), Article e27283.
Skjuve, M., & Brandtzæ g, P. B. (2018). Chatbots as a new user interface for providing health information to young people. *Youth and News in A Digital Media Environment–Nordic-Baltic Perspectives*.
Stanovich, K. E. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review*, *10*(1), 72–100.
Stencel, M. (2019). Number of fact-checking outlets surges to 188 in more than 60 countries.
Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge.
Strobelt, H., Oelke, D., Kwon, B. C., Schreck, T., & Pfister, H. (2015). Guidelines for effective usage of text highlighting techniques. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 489–498.
Sundar, S. S. (2008). *The MAIN Model: A heuristic approach to understanding technology effects on credibility*. MacArthur Foundation Digital Media and Learning Initiative Cambridge, MA.
Swire, B., Ecker, U. K., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1948.
Swire-Thompson, B., & Lazer, D. (2019). Public health and online misinformation: Challenges and recommendations. *Annual Review of Public Health*, *41*, 433–451.
Teo, T., & Zhou, M. (2014). Explaining the intention to use technology among university students: A structural equation modeling approach. *Journal of Computing in Higher Education*, *26*(2), 124–142.
Teoh, D. (2019). The power of social media for HPV vaccination–not fake news! *American Society of Clinical Oncology Educational Book*, *39*, 75–78.
Thomas, R. J., Tandoc Jr, E. C., & Hinnant, A. (2017). False balance in public health reporting? Michele Bachmann, the HPV vaccine, and "mental retardation". *Health Communication*, *32*(2), 152–160.
Tuch, A. N., Bargas-Avila, J. A., Opwis, K., & Wilhelm, F. H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies*, *67*(9), 703–715.
Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, *64*(7), 456–464.
Veglis, A., & Maniou, T. (2019). Chatbots on the rise: A new narrative in journalism. *Studies in Media and Communication*, *7*(1), 1–6.
Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.
Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, *39*(5), 621–645.
Walter, N., Brooks, J. J., Saucier, C. J., & Suresh, S. (2021). Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. *Health Communication*, *36*(13), 1776–1784.
Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, *85*(3), 423–441.
Walwema, J. (2021). The WHO health alert: Communicating a global pandemic with WhatsApp. *Journal of Business and Technical Communication*, *35*(1), 35–40.
Wintersieck, A. L. (2017). Debating the truth: The impact of fact-checking during electoral debates. *American Politics Research*, *45*(2), 304–331.
Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2018). Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck. org and FlackCheck. org. *Journalism & Mass Communication Quarterly*, *95*(1), 49–75.
Zarouali, B., Makhortykh, M., Bastian, M., & Araujo, T. (2021). Overcoming polarization with chatbot news? Investigating the impact of news content containing opposing views on agreement and credibility. *European Journal of Communication*, *36*(1), 53–68.
Zhang, J., Featherstone, J. D., Calabrese, C., & Wojcieszak, M. (2021). Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. *Preventive Medicine*, *145*, Article 106408.