

USING CHATBOTS TO CORRECT COVID-19 MISINFORMATION: A CONTROLLED
EXPERIMENT

BY

ZIYANG GONG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Advertising
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Master's Committee:

Assistant Professor Leona Yi-Fan Su, Adviser
Professor Kevin Wise
Professor Sela Sar

ABSTRACT

Conversational agents, otherwise known as chatbots, have been increasingly used by organizations worldwide to provide health-related information and correct misinformation in the COVID-19 pandemic. However, previous studies have neither empirically investigated the effectiveness of correcting misinformation using chatbots, nor have they provided theory-based guidelines for the designing of such chatbots. To address these concerns, we developed our misinformation-correcting chatbots and conducted two experiments. In the pre-test experiment (n = 42), we examined whether corrective information provided through chatbot interactions can correct COVID-related misperception and behavioral intention compared to not receiving such information. In the official study (n = 791), we investigate whether and how chatbots with interactive conversational cues and/or empathetic conversational cues can outperform the basic chatbot in correcting users' COVID-19-vaccination misperception and enhancing users' vaccination intention. Our results indicate that chatbots can be effective in fighting against misinformation and interactive and empathetic conversational cues do affect users' psychological perception towards the chatbot and in turn leads to better interaction outcomes. We also identified different mechanisms of people's misperception change and behavioral intention change. Theoretical and practical implications are discussed.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: PRE-TEST	4
CHAPTER 3: OFFICIAL STUDY.....	10
CHAPTER 4: DISCUSSION.....	25
CHAPTER 5: LIMITATION.....	31
CHAPTER 6: CONCLUSION	33
CHAPTER 7: FIGURES AND TABLES.....	35
REFERENCES	56
APPENDIX A: THE INSTRUCTION TO INTERACT WITH THE ASSIGNED CHATBOT ..	71

CHAPTER 1: INTRODUCTION

The astounding development of machine learning (ML) and natural language processing (NLP) in the past few decades has made conversational agents, which “accept natural language as input and generate natural language as output, engaging in a conversation with the user” (Griol et al., 2013, p. 760) increasingly prevalent. Amazon’s Alexa, Apple’s Siri, and Google Assistant are some of the most well-received voice-based conversational agents. They can assist users with shopping, provide daily news, entertain, and much more. Beyond the realm of smart speakers, conversational agents have been utilized in numerous fields, including education (Perez et al., 2020), medical consulting (Rosruen & Samanchuen, 2018), customer service (Cui et al., 2017), psychological counseling (Ho et al., 2018) and quite lately---fighting against misinformation (Amiri & Karahanna, 2022; Li et al., 2020; WHO, 2020).

Misinformation and disinformation in public health have been detrimental to people’s physical and mental health (Durizzo et al., 2021). It instigates discrimination and attacks towards infected people (Sotgiu & Dobler, 2020) and discourages observance of public health guidelines, thus conflicting with the efforts to contain the public health crisis (Hatcher, 2020). The long-existing and wide-ranging misinformation in public health have been highlighted in the COVID-19 pandemic. With conspiracy theories and misleading health advice being one of the defining characteristics of the pandemic, the World Health Organization (WHO) used the word “infodemic” to underscore the commensurate threats it might have compared to the virus itself (Zarocostas, 2020). To combat the “infodemic” and enhance the accessibility of reliable information, multiple organizations have leveraged on chatbots, including WHO (WHO, 2020), Poynter Institute’s International Fact-Checking Network (IFCN) (Grau, 2020), U.S. Department

of Veterans Affairs(USDVA, 2022), and many other institutes (Almalki & Azeez, 2020).

Almalki & Azeez (2020) reviewed the literature on COVID-19-related chatbots in healthcare and found there are currently five applications of these chatbots including (1) Disseminating health information, (2) Self-triage and risk assessment, (3) Monitoring and notifying exposure, (4) Tracking symptoms and health conditions (5) Correcting misinformation. The rising prevalence of these COVID-related chatbots indicates the promising potential it might have to disseminate reliable COVID-19 information and help fight misinformation.

However, as recognized by Miner et al (2020), despite chatbots' increasing popularity, the lack of research on their effectiveness is cause for concern (Miner et al., 2020). Besides, the designing of AI systems has also been accused of being based on "heuristic programming" instead of theories (Marcondes et al., 2020; Parnas, 2017). The present research aims to address these concerns by two controlled experiments. We trained chatbots to answer users' COVID-19-related questions phrased in various ways. This allows us to simulate realistic interaction with chatbots designed by other health organizations as opposed to instructing participants to simply read the transcripts of interactions (e.g., Liu & Sundar, 2018) or using "Wizard of Oz", in which participants are instructed to interact with a "chatbot", but it is actually researchers behind the screen (e.g., Ho et al., 2018). In our pre-test experiment, we examine the effectiveness of chatbots in correcting people's misperceptions and behavioral intentions. The findings of this research help clarify whether a chatbot could be an effective tool for combating misinformation. In the official study, we apply theoretical understandings of AI systems and people's misperception change on designing promising conversational features for chatbots. We investigate whether and how these features can enhance chatbots' efficiency in affecting people's COVID-19 vaccination misperception and vaccination intention. Our study is believed to be the

first to empirically examine chatbot's effectiveness in fighting against health misinformation and how chatbots' conversational features affect that potential, and as such, it is expected to provide insights to further understand people's misperception change and behavioral intention change mechanisms and help public health practitioners to design chatbots with the most effective features.

CHAPTER 2: PRE-TEST

A meta-analysis of the impact of attempts to correct health misinformation on social media found such attempts have a positive and significant mean effect size (Walter et al., 2021). However, chatbot, as a novel tool to fight against health-related misinformation, has not yet been empirically examined for its effectiveness in altering people's misperceptions. While corrective messages on social media might be provided by a peer, an expert, or a trusted organization, messages from a chatbot are provided by an algorithm. The existing study indicates that the chatbot's characteristic of providing algorithm-based responses might bring in a boost for its credibility due to the "automation bias" (Goddard et al., 2012), human's tendency to trust in automation. Bode & Vraga (2018) has investigated how Facebook related stories function can be used to correct Zika-virus-related misinformation. One of the related stories was produced by an algorithm and another is from other Facebook users. The result showed that algorithmic correction and social correction were equally effective in limiting people's misperception and the algorithmic correction was even perceived significantly more credible than social correction for those who report lower conspiracy beliefs. Since a chatbot is also a form of algorithm and embodies automation by replying to people's various inquiries, it should also be able to effectively correct participants' misperceptions.

Hypothesis 1: When providing corrective information regarding COVID-19, a chatbot can significantly decrease users' COVID-19-related misperception and their intentions to conduct dangerous behaviors against the novel coronavirus, compared to when such information was not provided.

2.1 PRE-TEST METHOD

2.1.1 Chatbot design

In the pre-test, we built our chatbot using Google’s Dialogflow, a natural language understanding platform that helps design and deploy chatbots. Previous academic studies have frequently used this tool in their chatbot design (Bozic et al., 2019; Reyes et al., 2019; Rosruen & Samanchuen, 2018). To develop our chatbot, we first collected training data via Amazon Mechanical Turk (MTurk). The collection plan was approved by the Institutional Review Board (IRB) of a large public university in the American Midwest. 30 workers were recruited and exposed to three questions related to COVID-19 (see Table 2). They were asked to imagine having similar concerns to those expressed in the questions and required to provide words, phrases, or sentences they would type into a chatbot to ask those questions. About 120 valid and varying inputs for each of the three questions were collected and started to become repetitive. We, then, used the collected inputs to train the chatbot. The training allows the chatbot to match inputs similar to those in the training material to one of the predefined answers. Inputs that fail to be recognized would trigger a default fallback intent¹ asking the user to only ask the questions they were required to ask. After the training, we deployed our chatbot through Dialogflow’s integration function to a website we built on Wix.com. See Figure 2 for a screenshot of the chatbot's interface for the pre-test.

¹ When the default fallback intent is triggered, the chatbot replies “Sorry, I am not currently able to answer this. Please make sure you have input your ID, and only ask the questions that you were provided within the survey.”

2.1.2 Procedure

Participants joined our study through an online survey hosted on Qualtrics. They were exposed to a consent form at the beginning of the survey indicating that the primary purpose of the study is to understand their media use and COVID-19-related attitudes and behaviors. After signing the consent form, they were exposed to a piece of misinformation related to COVID-19 without being informed that such information is false. The piece of misinformation claimed that the diluted bleach should be used to disinfect fruits and vegetables to prevent getting COVID-19 (see Figure 1). Then, questions were asked to evaluate their misperception and behavioral intentions entailed by the exposed misinformation. After answering several other questions, they were randomly divided into two groups and directed to a chatbot deployed on a website. Instructions on how to interact with it were given. In the control group, they were instructed to ask a chatbot two questions irrelevant to the misinformation (see Table 2). In the experimental group, participants were also required to ask the chatbot two questions, one of which is the same as the control group and the other is related to the misinformation. Participants were informed that the chatbot can understand questions phrased in various ways and they were instructed to input the questions provided in their own wording. After finished interacting with the chatbot, they returned to the survey. Questions evaluating their misperception and behavioral intention were asked again. Then, they completed a set of demographic questions. At the end of the survey, they were debriefed with the actual purpose of the study and given links to corrective and other relevant information related to COVID-19. Details of the questions and the chatbot's reply are summarized in Table 2. The wordings of key survey questions and the reliability scores are summarized in Table 3.

2.1.3 Participants

The official study plan was approved by the same IRB that approved our training data collection plan. 55 students were recruited through the University's Advertising Research Participation System (SONA system) in April and May 2021 and received course credit for their participation. Within the 55 participations, 2 reported currently not living in the US and 4 failed to complete the survey. They were excluded in further data analysis. Besides, after reviewing participants' chat history with the chatbot, 6 were found to have irrelevant input astray from our requirement. Another 1 asked the required question but the chatbot failed to understand the specific wording². Consequently, these 7 participants triggered the chatbot's default fallback response and thus, we excluded them to ensure all the participants in the further analysis experienced similar well-functioning chatbot interactions in line with our requirement. These excluded participants were not systematically attributed to the control or experiment group. The final sample contains 42 valid records with an average age of 21.0 ($SD = 1.01$), 83.3% of whom were female (see Table 1).

2.1.4 Manipulation checks

Two manipulation checks were utilized in the present study. One ensures participants read the misinformation carefully, and the other guarantees their interaction with the chatbot was valid. First, participants were asked to identify what COVID-19-related information was provided in the exposed figure, choosing from seven choices. 97.6 percent of the 42 participants

² The wording of this participant's input was notably distinct from the chatbot's training material. Thus, the chatbot failed to recognize the intent of that input. It is worthy to note that the inquiries about stress from other participants were all correctly answered by the chatbot, showing a 97.6% accuracy. For the other two questions, the chatbot has an accuracy of 100%. (Accuracy = times the chatbot responded correctly / times participants asked the required question)

successfully identified the topic of the exposed figure. Second, we asked participants to identify the information source mentioned in the chatbot's response. All participants correctly identified the source. These results ensure that participants carefully read the misinformation stimuli and chatbot's response.

2.2 PRE-TEST RESULTS

Independent t-tests were conducted to see whether the change of misperception and behavioral intention³ after interacting with a chatbot was significantly different between the control group and the experiment group. The 21 participants in the control group in which no corrective information was provided had a slight increase in terms of their misperception after interacting with the chatbot ($M = 0.095$, $SD = 0.199$), compared to the 21 in the experiment group in which the misperception was noticeably reduced ($M = -0.667$, $SD = 0.258$). The misperception change between the two groups was found significantly different, $t(40) = 2.3$, $p = .025$. Similarly, significant difference of the behavioral intention change between control group ($M = 0.206$, $SD = 0.183$) and experiment group ($M = -0.429$, $SD = 0.161$) was also observed $t(40) = 2.6$, $p = .013$. These results indicate that a chatbot when providing corrective information regarding COVID-19 can significantly decrease users' COVID-19-related misperception and their intentions to conduct dangerous behaviors against the novel coronavirus.

³ The change of misperception = misperception before chatbot interactions - misperception after chatbot interactions
The change of behavioral intention = behavioral intention before chatbot interactions - behavioral intention after chatbot interactions

2.3 PRE-TEST DISCUSSION

Our pre-test addressed the blank of literature by empirically demonstrating chatbots' potential in correcting people's misperception and mitigating their dangerous behavioral intentions. This finding is reassuring and consistent with the increasing efforts public health agencies put into developing chatbots to provide reliable health information (Amiri & Karahanna, 2022). We encourage more health-related organizations to incorporate chatbots as additional communication channels to make their information more accessible to their audience and help fight against health-related misinformation.

CHAPTER 3: OFFICIAL STUDY

Beyond demonstrating the effectiveness of chatbots in addressing misinformation, we want to further understand whether and how chatbots with different conversational cues meanwhile providing the same core corrective messages can have different efficiency in correcting people's misperception and behavioral intention. We specifically investigate whether and how *interactive conversational cues* and *empathetic conversational cues* can bring in unique psychological effects that contribute to efficient changes in users' misperception and behavioral intention. We choose to debunk the misinformation about mRNA COVID-19 vaccines altering human DNA and being harmful to the human body, a widespread myth identified and corrected by CDC (CDC, 2021). Being exposed to COVID-19-vaccine-related misinformation was found to positively associate with vaccine hesitancy (Biswas et al., 2021) which can put individuals at higher risk of getting and spreading the virus that causes COVID-19 and developing serious illness and death (CDC, 2022).

3.1 CHATBOT ARE SOCIAL ACTORS

Many previous chatbot studies examine how certain chatbot features can lead to different, preferably enhanced, conversational outcomes (e.g., Araujo, 2018; Sundar et al., 2016; Toader et al., 2020). This line of research contributes insights for practitioners to maximize the conversational impact by optimizing the chatbot design. Computers are social actors (CASA; Nass et al., 1994; Nass & Moon, 2000) is a theoretical framework frequently employed to guide these studies (Gambino et al., 2020). It suggests that humans treat *media agents* (i.e.,

“technological artifact that demonstrates sufficient social cues to indicate the potential to be a source of social interaction”, Gambino et al., 2020, p. 73) similarly to how they would interact with another real human being, despite knowing those agents are inanimate (Nass & Moon, 2000). In other words, instead of questioning media agents’ social behavior, humans tend to respond mindlessly to them to the extent that they applied their existing mental model learned from human-human interaction to the human-agent interaction. Existing literature constructed media agents’ human-like traits and their potential for human-like social interaction as *anthropomorphism* (e.g., Breazeal, 2003; Waytz et al., 2010). This concept has been regarded as a critical factor in the evaluation of media agents (Gambino et al., 2020). An extensive literature review of the acceptance of social robots done by de Graaf & Ben Allouch (2013) found that anthropomorphism is positively associated with many outcomes such as attractiveness, sociability, companionship, and attitude towards robots. More specifically, empirical studies found that it is through *social presence*, a feeling that the media agent is a social entity *being with* the users (Kilteni et al., 2012), that chatbots’ anthropomorphic features exert their influence on the outcomes of human-chatbot interactions such as customers’ compliance with chatbot’s requests (Adam et al., 2021), the satisfaction of a service (e.g., Verhagen et al., 2014), perceived expertise, perceived friendliness, attitudes, and behavioral intention towards the chatbot (Go & Sundar, 2019).

Conversational cues (i.e., human-like languages) were regarded as one of the three types of anthropomorphic features of chatbots (Go & Sundar, 2019). They were found to be able to compensate for the inanimate nature resulting from less anthropomorphic visual cues and identity cues and have the potential to bring in unique psychological effects that contribute to positive attitudinal and behavioral outcomes (Go & Sundar, 2019). In the official study, we

investigate two types of conversational cues, *interactive conversational cues* and *empathetic conversational cues* for their potential to improve chatbot's impact in affecting users' misperception, behavioral intention.

3.2 INTERACTIVE CONVERSATIONAL CUES

One of the most remarkable natures that set chatbot aside from misinformation-correcting social media posts and fact-checking websites is its *interactivity*. Although there has not been a universally accepted definition of interactivity, several elements have been summarized by Sundar et al. (2016) including the ability to influence the content (e.g., Jensen, 1998), personalization (e.g., Wu, 2006), reciprocal communication (e.g., Jiang et al., 2010), and user's control over media (e.g., Coursaris & Sung, 2012). According to the Modality-Agency-Interactivity-Navigability model (MAIN, Sundar, 2008), interactivity is one of the most essential technological affordances (i.e., actionable structural features, Norman, 1999) that could affect the perceived credibility of digital media. A high level of interactivity in chatbots might trigger several types of heuristics which in turn could affect the perceived credibility of audiences. For example, allowing users to input customized questions instead of having them choose from a list of given questions might serve as a cue to trigger the *responsiveness heuristic* which indicates to the user that the chatbot is capable of considering the variation in user input and accordingly offering dynamic customized content (Sundar, 2008). Users who have this heuristic may tend to believe the chatbot is of good quality, which eventually enhances the credibility of the chatbot (Sundar, 2008). Besides, interactivity operationalized by enabling users to customize the order of information to be present (Skalski & Tamborini, 2007), and by allowing users to choose from

different conversational agents and their characteristics (Xiao et al., 2007) might also trigger another heuristic---*control heuristic*, which would render a sense that the user is controlling the nature of their interaction and is managing the pace and direction of information acquisitions (Sundar, 2008). This heuristic might help to exalt the perceived information value and consequently enhance its credibility (Sundar, 2008).

In the context of human-computer interaction (HCI), a key characteristic of interactivity, contingency, has gained much attention (e.g., Bellur & Sundar, 2017; Go & Sundar, 2019; Ischen et al., 2020; Sundar et al., 2016). This concept is also known as message interactivity (or semantic interactivity, Sohn, 2011), which is defined as “affordances that enable users to interact with the system and/or with another user in a contingent manner to achieve a sense of dialogue” (Sundar et al., 2016, p. 600). Go & Sundar (2019) operationalized message interactivity as conversational cues with or without back-and-forth message exchanges. In their study, chatbots with high message interactivity address the previous user’s inputs before each response, while chatbots with low message interactivity directly respond to the user’s inputs without mentioning the previous dialog. In line with the MAIN model, they found that the agent’s high message interactivity leads to a higher level of social presence which enhanced the *perceived expertise* of the chatbot, a sub-dimension of *credibility* (Corritore et al., 2003), and *perceived trustworthiness* (Mayer et al., 1995).

Besides having enhanced credibility, interactive features also have the potential to enhance users’ attention towards the content presented (e.g., Chua & Banerjee, 2015; Geri et al., 2017; Xu & Sundar, 2016). Although there haven’t been enough studies specifically examining the relationship between chatbot’s interactive features and user’s attention (but see Sundar et al., 2016), interactive features in other domains demonstrated their positive effects. Geri et al.

(2017), for example, found interactive elements in video lectures can expand learners' attention span and increase the completion percentage of video-watching. In the context of e-commerce websites, interactive features were also found to attract users' attention and engage them to process the information presented (Xu & Sundar, 2016). Chua & Banerjee (2015) also identified interactivity as one of the three factors that draw the audience's attention to businesses' social media posts.

Literature in correcting misinformation pointed out that the credibility of the source of the corrective information (e.g., Kim et al., 2021; Walter & Murphy, 2018) as well as audiences' attention when being exposed to the corrective information (e.g., Kim et al., 2021) plays important and positive roles in correcting misperception. Given the potential benefits of enhanced credibility and attention induced by interactive features, highly interactive chatbots might outperform lowly interactive chatbots in terms of their efficiency in correcting misinformation.

Aligned with Sundar et al. (2016), we consider users' perceived interactivity to be critical in bringing in the positive effects of interactive features we manipulate. We adopt their definition of perceived interactivity as "a users' subjective and overall assessment of interactivity" (p. 601) which includes *perceived contingency*, a concept specifically "operates at the level of threadedness and interconnectedness of message exchange" (p. 601). Examining users' perceived interactivity not only helps validate our manipulation of interactivity but also increases the generalizability of our findings to situations where the interactive features were not operationalized exactly the same as our manipulation. Accordingly, we hypothesize:

Hypothesis: (H2) a misinformation-correcting chatbot with interactive conversational cues can lead to a higher level of perceived interactivity which can induce (H3) a larger *misperception change* and (H4) a larger *change in behavioral intention* compared to a chatbot without such cues

3.3 EMPATHETIC CONVERSATIONAL CUES

Empathy is defined as the ability to understand and identify others' mental states and share the feeling of others (Kerr-Gaffney et al., 2019). It was regarded as a multifaceted concept consisting of cognitive empathy as well as affective empathy (e.g., Campbell & Babrow, 2004; Pérez-Fuentes et al., 2020; Shechtman, 2002). The former denotes the comprehension of others' emotions while the latter emphasizes feeling that same emotion (Vossen et al., 2015). While existing literature calls for more attention in addressing emotional factors such as showing empathy and compassion when correcting misinformation (e.g., Gesser-Edelsburg et al., 2018; Hardee, 2020; MacKay et al., 2021), empirical findings regarding how empathy affects misperception correction are less consistent (e.g., Gesser-Edelsburg et al., 2018; Tully et al., 2020). Gesser-Edelsburg et al. (2018) found that when addressing misinformation about measles vaccination, corrective information that is transparent, showing empathy, and referring to the audience's fears and concerns can lead to higher perceived satisfaction compared to an unequivocal corrective message without addressing the emotional element. They also found a significant association between that satisfaction and behavioral intention to vaccinate children. Tully et al. (2020), however, found whether or not acknowledging and affirming the audience's

confusion and uncertainty when correcting food-related misinformation does not make corrective messages more effective in reducing misperception.

Empathy has also been examined in conversational agents in many contexts such as medical consulting (e.g., Liu & Sundar, 2018), mental health consulting (e.g., Harilal et al., 2020), and intervening drinking behaviors (e.g., Lisetti et al., 2013). These empirical studies indicate a neutral-to-positive effect of empathetic conversational cues depending on what specific conversational outcomes are examined. For example, Kraus et al. (2021) investigated chatbot in the mental health domain and found small talks with empathetic conversational cues does not significantly foster human-chatbot trust. Liu & Sundar (2018), however, found that, in the context of providing advice for sexually transmitted diseases (STD), chatbot's empathetic expression was favored over unemotional expression. Lisetti et al. (2013) compared three types of conversational agents (text-based, empathetic embodied agent, non-empathetic embodied agent) in providing drinking problem intervention and found the empathetic embodied one has many benefits over others including better acceptability, enjoyability, trustworthiness, and intelligence.

Different from these empirical findings, another popular theoretical framework, the Uncanny Valley of Mind (UVM), suggests that when the “digital mind” demonstrate their abilities related to human distinctiveness such as empathy and social cognition, people will feel threatened by this violation of human distinctiveness and perceive that “digital mind” as eerie (e.g., Ferrari et al., 2016; Liu & Sundar, 2018; Stein & Ohler, 2017). However, several literature reviews conclude that empirical evidence for the hypotheses of UVM was inconsistent (e.g., Kätsyri et al., 2015; Zhang et al., 2020). In the context of chatbots providing health-related information, Liu & Sundar (2018) tested UVM against its contradictory framework, CASA.

They found CASA to be more applicable than UVM for people who disbelieve robotic socioemotional capabilities. More studies need to be done to see whether the two frameworks have different applicability under different contexts. Given the inconsistent findings of the impact of empathy both in misinformation- and chatbot-related literature, we propose the following two research questions and one hypothesis:

H5: a misinformation-correcting chatbot with empathetic conversational cues can lead to a higher level of *perceived empathy* compared to the chatbot without such cues.

Research Questions: Whether and how chatbot's perceived empathy mediates user's (RQ1) *misperception change* and (RQ2) *change in behavioral intention*.

3.4 OFFICIAL STUDY METHOD

3.4.1 Operationalize interactive/ empathetic conversational cues

In line with previous literature (e.g., Go & Sundar, 2019; Lew et al., 2018), the present study also operationalizes interactivity as being able to address the user's question before answering it. In addition to that, the interactive chatbot in our study will deliver the answer not just through a single scripted response but through back-and-forth contingent dialogues. After the chatbot provides a brief answer, it asks the user whether more information is needed. More information will be provided once the user reponed affirmatively. In contrast, the non-interactive chatbots will not address the user's input and will provide the same information altogether without embedding it through active and two-way dialogue. Allowing users to accept/decline

additional corrective information inevitably split users into those who received the full amount of corrective information and those who received only partial corrective information. In the present study, we report the findings regarding those who received only partial corrective information as additional findings.

We operationalize empathy based on Liu & Sundar (2018). They use sentences that acknowledge the user's feelings to operationalize *cognitive empathy* and use sentences that express the chatbot's understanding of how and why the user feels a certain way as *affective empathy*. We incorporate the two sub-dimensions of empathy as one in the present study. (See Table 4 for details of the operationalization).

3.4.2 Chatbot design

The chatbots in the official study were built and trained in the same way as in our first study. A total of 101 valid and varying training phrases were collected from MTurk. Next, we integrated the four Dialogflow chatbots with Kommunicate.io, a platform that assists customization of the chatbot interface, and deployed them on a website we build on Wix.com. See Figure 3 for screenshots of chatbots' interface.

3.4.3 Participants

A total of 1155 participants based in the U.S. and who haven't received a coronavirus vaccination at all were recruited from 13 Oct 2021 to 9 Nov 2021 through Prolific, a crowdsourcing platform frequently used by industry and academia. Previous research compared samples recruited from crowdsourcing platforms such as Amazon Mechanical Turk (MTurk) with population-based ones and found that their respective estimations of treatment effects were

comparable and that the MTurk samples' generalizability was acceptable (Mullinix et al., 2015). Another study compared the different crowdsourcing platforms and found that Prolific tends to have higher data quality than MTurk (Peer et al., 2021). All recruited Prolific participants were required to complete the survey using a desktop computer. Each participant was paid a fee of \$2.22 for their efforts.

To ensure a smooth and comparable user experience when interacting with the chatbot, specific instruction was given to guild participants (see appendix 1). After reviewing participants' chat history with the chatbots, we found 337 of them failed to follow the instruction while interacting with the chatbot. These issues include (1) interacting with the chatbot multiple times; (2) asking the required question more than once; (3) forgetting to input the ids assigned in the survey to initiate the chatbot in the beginning; (4) asking irrelevant questions; (5) making unrequired social small talk such as "Have a nice day computer"; (6) making unrequired comments on the information provided by the chatbot such as "This sounds like propaganda"; (7) making demanding input such as "next", "end", "finished". Since these inputs were astray from our instruction, which the chatbots were not trained to respond to, they might lead to notably less fluent interactions with the chatbot. Participants might perceive the chatbot as having lower quality compared to those who did not encounter these issues. Therefore, they are excluded from further analysis. Other 27 participants followed the instruction but encountered technical issues. 24 of them were not correctly assigned to their designated chatbot.⁴ The other 3 have their inputs as required but in a distinctively different way from our training database so the chatbot failed to understand. They were also excluded from further analysis. The final sample contains 791 valid records.

⁴ This issue appears to result from a bad internet connection at the user's side so that the web page fails to load correctly.

3.4.4 Procedure

Participants joined our study through an online survey hosted on Qualtrics. They were exposed to a consent form at the beginning of the survey indicating that the primary purpose of the study is to understand their media use and COVID-19-related attitudes and behaviors. After signing the consent form, they were exposed to a piece of misinformation related to COVID-19 vaccines without being informed that such information is false. That misinformation claimed that injecting mRNA COVID-19 vaccines might alter human DNA. (See Figure 4). Then, questions were asked to evaluate their misperception and vaccination intentions entailed by the exposed misinformation. After answering several other questions, they were introduced to what a conversational agent is and shown an example built by WHO (see Figure 5). Questions were asked to evaluate their usage of chatbots in everyday life. Then, they were divided into four groups and directed to one of four chatbots deployed on a website including (1) the basic chatbot, (2) the empathetic chatbot, (3) the interactive chatbot, and (4) the interactive & empathetic chatbot. The chatbots with the interactive feature allow participants to choose whether to receive full corrective information. If users choose not to receive it, the amount of corrective information they received will be less than the other two conditions (See Figure 3). Since we are interested in comparing chatbots with different conversational cues but provide the same corrective information, we assigned more participants to the two interactive chatbot conditions so that the number of participants who choose to receive full information roughly equals the number of participants in the basic chatbot condition and empathetic chatbot condition. As a result, 146 interacted with a basic chatbot (BC), 142 interacted with the empathetic chatbot (EC), 166 interacted with the interactive chatbot and choose to receive full

information (IC); 156 interacted with the interactive & empathetic chatbot and choose to receive full information (I&EC); 100 interacted with the interactive chatbot but choose to receive partial information (IC-partial); 81 interacted with the interactive & empathetic chatbot but choose to receive partial information (I&EC-partial). After finished interacting with the chatbot, they returned to the survey. Questions evaluating their misperception and vaccination intention were asked again. Their cognitive and affective perceptions of the interactions with the assigned chatbot were also measured. Then, they completed a set of demographic questions. At the end of the survey, they were debriefed with the actual purpose of the study and given links to corrective and other relevant information related to COVID-19. The wordings of key survey questions and the reliability scores are summarized in Table 6.

3.4.5 Manipulation checks

Besides checking participants' chatting history with the chatbot, other two manipulation checks were also utilized in the official study. The first one examines whether participants understand the shown misinformation and the other one ensures participants were not significantly different among groups. First, participants were asked what COVID-19-related information was provided after exposing them to the misinformation. 98.9% of the 791 valid participants correctly selected the topic of the exposed information. Second, Chi-square and ANOVA were conducted to variables including gender, age, household income, education, race, chatbot usage. The results found no significant difference among the six types of interaction in terms of any of these variables (See Table 5).

3.4.6 Data Analysis

We conducted data analysis using IBM SPSS Statistics (Version 26.0). To test our Hypotheses and research questions, we conducted ordinary least squares (OLS) regressions using the SPSS version of the PROCESS macro (A. Hayes, 2017), a modeling tool widely used in moderation and mediation analysis (e.g., Ahorsu et al., 2020; Emery et al., 2017; Opie et al., 2020). Adopting PROCESS's pre-programmed Model 4, we specified our independent variable, i.e., the six types of interactions, as multi-categorical, and computed confidence intervals with 10,000 bootstrap samples. To avoid violating the assumption of homoskedasticity in OLS regression, we used heteroskedasticity-consistent standard error estimators, HC3, for OLS regression as recommended by previous studies (e.g., A. F. Hayes & Cai, 2007; Long & Ervin, 2000).

3.5 OFFICIAL STUDY RESULTS

3.5.1 Hypotheses 2, 3, and 4

OLS regression found that compared to BC, EC (Table 7; $B = 0.35$, $SE = 0.17$, $p = .035$)⁵, IC (Table 7; $B = 0.53$, $SE = 0.15$, $p < .001$), I&EC (Table 7; $B = 0.67$, $SE = 0.16$, $p < .001$), IC-partial (Table 7; $B = 0.52$, $SE = 0.18$, $p = .003$), and I&EC-partial (Table 7; $B = 0.59$, $SE = 0.21$, $p = .005$) all significantly raised the perceived interactivity of the chatbots. Thus, H2 was supported.

⁵ Our results show that the chatbot designed with empathetic features but not with interactive features still leads to a higher level of perceived interactivity compared to the basic chatbot. This is because acknowledging users' feelings using expressions such as "I understand your worries" inevitably achieve a sense of active and engaging dialogue which is closely related to the definition of message interactivity (Sundar et al., 2016)

As expected, perceived interactivity was positively associated with misperception change (Table 7; $B = 0.17$, $SE = 0.04$, $p < .001$) and change in vaccination intention (Table 7; $B = 0.08$, $SE = 0.02$, $p = .001$). Bootstrap estimation showed significant indices for all these mediation pathways from chatbots conditions via perceived interactivity to the two outcome variables--- misperception change and vaccination intention (See Table 8). Thus, H3 and H4 were supported.

3.5.2 H5, RQ1, and RQ2

OLS regression also found that compared to BC, EC (Table 7; $B = 1.23$, $SE = 0.17$, $p < .001$), I&EC (Table 7; $B = 1.27$, $SE = 0.17$, $p < .001$), and I&EC-partial (Table 7; $B = 0.67$, $SE = 0.23$, $p = .004$) significantly raised the perceived empathy of the chatbots while IC (Table 7; $B = 0.29$, $SE = 0.18$, $p = .106$) IC-partial (Table 7; $B = 0.05$, $SE = 0.20$, $p = .790$) did not have significant effect on the perceived empathy. Therefore, H5 was supported.

Perceived empathy was found to be positively associated with change in vaccination intention (Table 7; $B = 0.08$, $SE = 0.02$, $p = .001$) but not for misperception change (Table 7; $B = 0.03$, $SE = 0.03$, $p = .342$). Bootstrap estimation showed non-significant indices for mediation pathways from chatbots conditions via perceived empathy to misperception change. Significant indices were found for mediation pathways from the three chatbots conditions that have empathetic conversational cues (i.e., EC, I&EC, and I&EC-partial) via perceived empathy to vaccination intention (See Table 8). The coefficients of each path for misperception change and vaccination intention are summarized in Figure 6 and Figure 7.

3.5.3 Additional findings

Our results indicates that when users refused to receive additional information from the chatbot, i.e., IC-partial condition (Table 7; $B = -0.61$, $SE = 0.11$, $p < .001$) and I&EC-partial condition (Table 7; $B = -0.50$, $SE = 0.13$, $p < .001$), there will be negative direct effects of the conditions to the *misinformation change* compared to BC. As for the *vaccination intention*, only IC-partial condition was found to have a significant and negative direct effect (Table 7; $B = -0.21$, $SE = 0.08$, $p = .005$) whereas I&EC-partial condition did not have a significant direct impact (Table 7; $B = -0.17$, $SE = 0.10$, $p = .094$).

CHAPTER 4: DISCUSSION

Taken together, our two experiments provide two main contributions. First, we empirically demonstrated chatbot's impact in correcting people's misperception and dangerous behavioral intentions. Second, our results revealed how chatbots' conversational cues can further boost such impact. It is worthy to emphasize that even when the core corrective messages provided by the chatbots were all identical, chatbots that were built with certain conversational features appear to affect users' psychological perception towards the chatbot which leads to better interaction outcomes than the chatbot without such features. In particular, we found that when answering people's inquiries about whether COVID-19 mRNA vaccines can alter human DNA and be harmful to the human body, chatbots with interactive cues can enhance users' perceived interactivity towards the chatbot and in turn, resulted in enhanced efficiency in *reducing misperception and increasing vaccination intention* compared to a basic chatbot that does not have interactive nor empathetic conversational cues. Chatbots with empathetic cues, on the other hand, can lead to a higher level of perceived empathy, which improves the efficiency in increasing vaccination intention but did not help reduce the misperception.

4.1 THE MEDITATION EFFECTS OF PERCEIVED INTERACTIVITY

Extending previous literature about *interactive features* in HCI (e.g., Go & Sundar, 2019; Ischen et al., 2020), our study further demonstrated its advantages in the context of addressing health-related misinformation using chatbots. Consistent with our hypothesis, our data support that even when the corrective information is identical, providing it through an interactive

dialogue rather than a once-and-for-all script did affect people's perception of the chatbot and eventually lead to better *misperception change* and *vaccination intention*. What's exciting about our finding is that interactive conversational cues can be easily embedded to chatbots by simply greeting, addressing the user's question before answering it, and giving the user more autonomy in the conversation and can, in a large scale, enhance the outcomes of the corrective information without a need to change the content of the corrective messages.

However, providing corrective messages through an interactive conversation inevitably enables users to refuse the offering of the full amount of correction. It might be a double-edged sword since receiving partial correction was found to lead to less misperception change. Further studies need to be done to empirically compare the pros and cons of breaking down an elaborated chatbot answer into multiple pieces and allowing users to choose the quantity of information they want. Other than that, we do recommend public health practitioners increase the interactivity of their chatbots so that users will have higher perceived interactivity. To do so, practitioners can try to enable chatbots to greet the users in the beginning and address users' inputs before answering them.

4.2 THE MEDITATION EFFECTS OF PERCEIVED EMPATHY

The empathetic conversational cues embedded in chatbots seem not to trigger an uncomfortable and eerie feeling as suggested by UVM. Instead, these cues affected the perceived empathy of the chatbot, which positively mediates empathetic conversational cues' effect on vaccination intention. Consistent with the previous study that used health chatbots to address

STD (Liu & Sundar, 2018), these findings also indicate higher applicability of the CASA paradigm rather than the UVM paradigm when it addresses vaccine-related misinformation.

The beneficial effect of perceived empathy on vaccination intention can be explained by the Integrated Model of Advice Giving (IMAG, Feng, 2009, 2014) which suggest that helping distressed other to overcome their emotional upsets should be an important initial step before problem inquiry and advice-giving because individuals might not be ready to closely look into the problem they are facing until they work through their emotional state. Feng (2009, 2014) found that not only the presence of the emotional support but also the sequential placement of the emotional support matters. Specifically, advice after providing emotional support and problem inquiry and analysis was judged to be higher in quality and was more likely to be followed compared to giving advice following other sequential patterns. For our participants, being exposed to misinformation that claims the mRNA vaccine is unsafe and under-researched might put them into a negative emotional state. In this case, the empathetic conversational cues might have helped acknowledge and validate their emotion, making them more emotionally ready to reappraisals the upsetting misinformation and take the corrective information as advice.

4.3 DIFFERENT MECHANISMS FOR THE CHANGE OF MISPERCEPTION AND VACCINATION INTENTION

Although knowledge and positive perceptions towards COVID-19 vaccination were previously found to predict better attitudes towards COVID-19 vaccination and in turn, associated with higher vaccination intention (Fan et al., 2021). Good COVID-19-related knowledge did not always be the precursor of preventative behaviors and vice versa (Nwagbara

et al., 2021). Our finding extends the literature by identifying the different mechanism of the people's misperception change and vaccination intention change. Both perceived interactivity and perceived empathy play positive roles in people's vaccination intention change but only perceived interactivity is positively associated with people's misperception change. This might be due to the different nature of the two outcomes investigated. Whereas vaccination intention is a preventative *behavioral intention* for COVID-19, misperception is considered as an incorrect perception of reality (Thampi, 2017). While people's learning process can be boosted by interactive systems (e.g., Evans & Gibbons, 2007) which help enhance their knowledge and correct the distorted perception of facts, our findings indicate that such process cannot be facilitated by simply switching to an empathetic tone. As for behavioral intention, however, extant studies pointed out the important role of emotion in people's planned behaviors (e.g., Ajzen, 2011; Y. J. Kim et al., 2013; Londono et al., 2017; Parkinson et al., 2018). People's anticipated negative emotions were found to decrease people's behavioral intention (e.g., Kim et al., 2013; Londono et al., 2017). This is consistent with our findings. Chatbot's empathetic expression might have curbed participants' anticipated negative emotion towards COVID-19 vaccination which helped boost their intention to take the vaccine. Empirical studies examining misperception correction also show that when looking at the behavioral intention as the outcome, empathetic corrective messages were found to lead to better satisfaction than a message without addressing the emotional element, which is positively associated with behavioral intention to vaccinate children (Gesser-Edelsburg et al., 2018). However, when focusing on the misperception as the outcome, whether the tone of the corrective messages was affirming, neutral, or uncivil does not matter anymore (Bode et al., 2020). Bode et al., (2020) further suggested that social media users' efforts to correct misinformation can be in whatever tone they feel most

comfortable. Our findings, however, suggest that although empathetic tones fail to boost people's misperception change, it does foster behavioral intention change. We encourage future efforts to correct misinformation to adopt empathetic tones and address the audience's emotion so that the core corrective message can be more effective to lead to positive behavioral changes.

4.4 DISCUSSION ABOUT ADDITIONAL FINDINGS

Allowing users to accept/decline additional corrective information inevitably split users into those who received the full amount of corrective information (i.e., IC and I&EC) and those who received only partial corrective information (i.e., IC-partial and I&EC-partial). Our results indicate that compared to BC condition who received the full corrective information, IC-partial and I&EC-partial have a direct negative effect on users' misperception change and IC-partial also have a direct negative effect on vaccination intention. This is in line with existing literature (e.g., van der Meer & Jin, 2020). Van der Meer & Jin (2020) also found that compared to participants who received simple corrective information, those exposed to more detailed and elaborated corrective information are better corrected and more likely to take preventive actions. What's surprising is that though IC-partial has a negative direct effect on vaccination intention, I&EC-partial does not. This indicates that, by adding empathetic conversational cues to the chatbots, the chatbots seem to be able to mitigate the adverse effect of receiving only partial corrective information on the vaccination intention. This can be explained by the feelings-as-information theory (FIT, Schwarz, 2012) which suggests that people use their subjective feelings such as emotions, moods, and easy or difficult processing experiences as a source of information in their judgment. FIT proposes these feelings play important roles in persuasion and processing

style that a positive feeling can lead to a less effortful and heuristic processing style whereas a negative feeling tends to result in a more bottom-up and systematic processing style (Schwarz, 2012). They further suggested that when people received persuasive information, recipients with negative feelings can be persuaded by strong arguments but less likely by weak arguments, and recipients with positive feelings can be equally persuaded by both kinds of arguments (Schwarz, 2012). When chatbot users only requested and received partial corrective information the arguments might be perceived as weaker compared to full corrective information. The reason that the I&EC-partial condition did not have a negative direct effect on vaccination intention as IC-partial did might be that the empathetic conversational cues initiated a positive emotion towards the chatbot which affected users' processing style and enhanced their receptivity to a relatively weaker argument. Whereas in the IC-partial condition, users might have a relatively not-so-positive emotion compared to I&EC-partial, which makes them stick to a systematic processing style and be more suspicious to a relatively weaker argument.

CHAPTER 5: LIMITATION

The present studies also have several limitations. Firstly, our study recruited people who has not been vaccinated with a covid-19 vaccine at the time of recruitment, which is inevitability unrepresentative. However, we argue that this sample captured whether and how misperception can be changed by chatbots' conversational cues for those who very likely have a certain level of initial misperception towards the investigated topic. Future studies can further investigate how chatbots' conversational cues work for different groups of people with different beliefs.

Secondly, our chatbots did not provide perfect replies to every user's input, especially when users' inputs are irrelevant to what we required in the instruction. Since the way participants deviate from our instruction is vastly different, we did not apply a theory-driven guideline to determine what types of deviation should be excluded or included. To prevent participants who encountered issues in their chatbot interactions from biasing our results, we excluded all the 364 participants who deviated from our instruction, which unavoidably brought some biases to our experiment. Nevertheless, our manipulation checks indicate that after excluding them no significant biases were identified across our experimental conditions in terms of sex, age, race, income, education, and previous chatbot usage. Besides, our post hoc analysis confirmed that when setting the demographic variables (i.e., age, gender, and race) as covariates, the model result will be similar and lead to the same interpretation. Thirdly, we only investigated the effects of conversational cues in the topic of COVID-19-vaccination, a topic that usually involves negative emotions (e.g., Alamoodi et al., 2021; Chou & Budenz, 2020). It is yet to be seen whether the perceived empathy will still exert its strength in affecting people's behavioral intention for other less emotional and controversial topics. Fourthly, we measured the

misperception and behavioral intention twice, once after exposing participants to misinformation and another time after participants interacted with the chatbot. This design might make some of the participants guess the real intention of our study and deviate from how they would report if they do not know our intention. Nevertheless, this likelihood should not be systematically biased between our experiment groups as they were measured twice in all groups. Our between-group findings should not be dampened by this design. Fifthly, our participants were directly exposed to the misinformation before interacting with the chatbot which is somewhat different from how people usually encounter misinformation and develop misperception. In this case, people might have different perceptions of the misinformation card and might have different levels of misperception toward the investigated topics after viewing the card. However, we argue that such differences should not be systematically biased across different experiment groups nether as all groups viewed the same cards. Lastly, misinformation was previously found to have continued influences that affect people's memory and reasoning even when the correction is well remembered (e.g., Ecker et al., 2010; Walter & Tukachinsky, 2020). Our studies have not investigated whether chatbots can better reduce the continued influence than other misinformation-correcting tools. We encourage future studies to investigate this in a longitudinal study to see whether chatbots have such potential. Despite having limitations, neither of them should have seriously dampened our experiment results and we argue that our findings still provide important theoretical and practical implications for researchers and practitioners.

CHAPTER 6: CONCLUSION

Chatbots are becoming increasingly popular in combating the novel SARS-CoV-2 coronavirus and its related misinformation. Despite the popularity, existing studies about COVID-19-related studies mainly focused on chatbots' development, architecture, and users experience (Almalki & Azeez, 2020). Little attention has been spent on examining the effectiveness of chatbots in correcting people's COVID-19-related misperception and behavioral intention. Our study addressed this gap by empirically demonstrating that when corrective information is received from a chatbot, participants' misperception and behavioral intention can both be corrected compared to not receiving such information. This suggests that apart from existing misinformation-correction tools such as fact-checking websites and social media, chatbots can also be an effective tool to fight against misinformation. Moreover, the designing of AI systems has been accused of mostly based on programmers' heuristics instead of theories (Marcondes et al., 2020; Parnas, 2017). Our studies address this concern and provide some guidelines for the designing of misinformation-correction chatbots. Specifically, we emphasize the beneficial effects of interactive conversational cues and empathetic conversational cues as they can affect users' psychological perception of the chatbot and in turn, have enhanced efficiency in correcting users' misperception and behavioral intention. We also identified different mechanisms of users' misperception change and behavioral intention change. In particular, we revealed that though perceived empathy did not exert positive effect on misperception change, it did foresteer behavioral intention. This highlighted the importance of tone when correcting misinformation. Although adopting an empathetic tone in the corrective messages might not affect people's misperception change, it appears to make it easier for people

to internalize the information and change their behavior. Based on these findings, we encourage public health organizations to incorporate chatbots as their additional channel to provide reliable health-related information to help fight against misinformation, and we recommend them to include interactive and empathetic conversational cues to their chatbots so their effectiveness of correcting misperception and behavioral intention can be enhanced.

CHAPTER 7: FIGURES AND TABLES

Figure 1: Misinformation card for the pre-test: disinfecting fresh produce with diluted bleach is recommended.

FACT:

To protect yourself against COVID-19, washing your fruits and vegetables with just water might not be enough. The novel coronavirus could multiply on these foods' surfaces and remain dangerous when eaten unless you disinfect them.

It is recommended to use diluted bleach to inactivate the virus. Make sure the bleach solution is diluted to about 30% concentration, and rinse fresh fruits and vegetables for more than 10 seconds before serving.

Disinfecting fruits and vegetables with diluted bleach is highly recommended to prevent you from getting COVID-19

#COVID19 #coronavirus


An illustration at the bottom center of the card shows a white bowl filled with fresh produce, including a purple grape, a red apple, a green leafy vegetable, and a carrot. To the right of the bowl is a white bottle with a brown label that says "Bleach".

Figure 2: A screenshot of the chatbot's interface in the pre-test

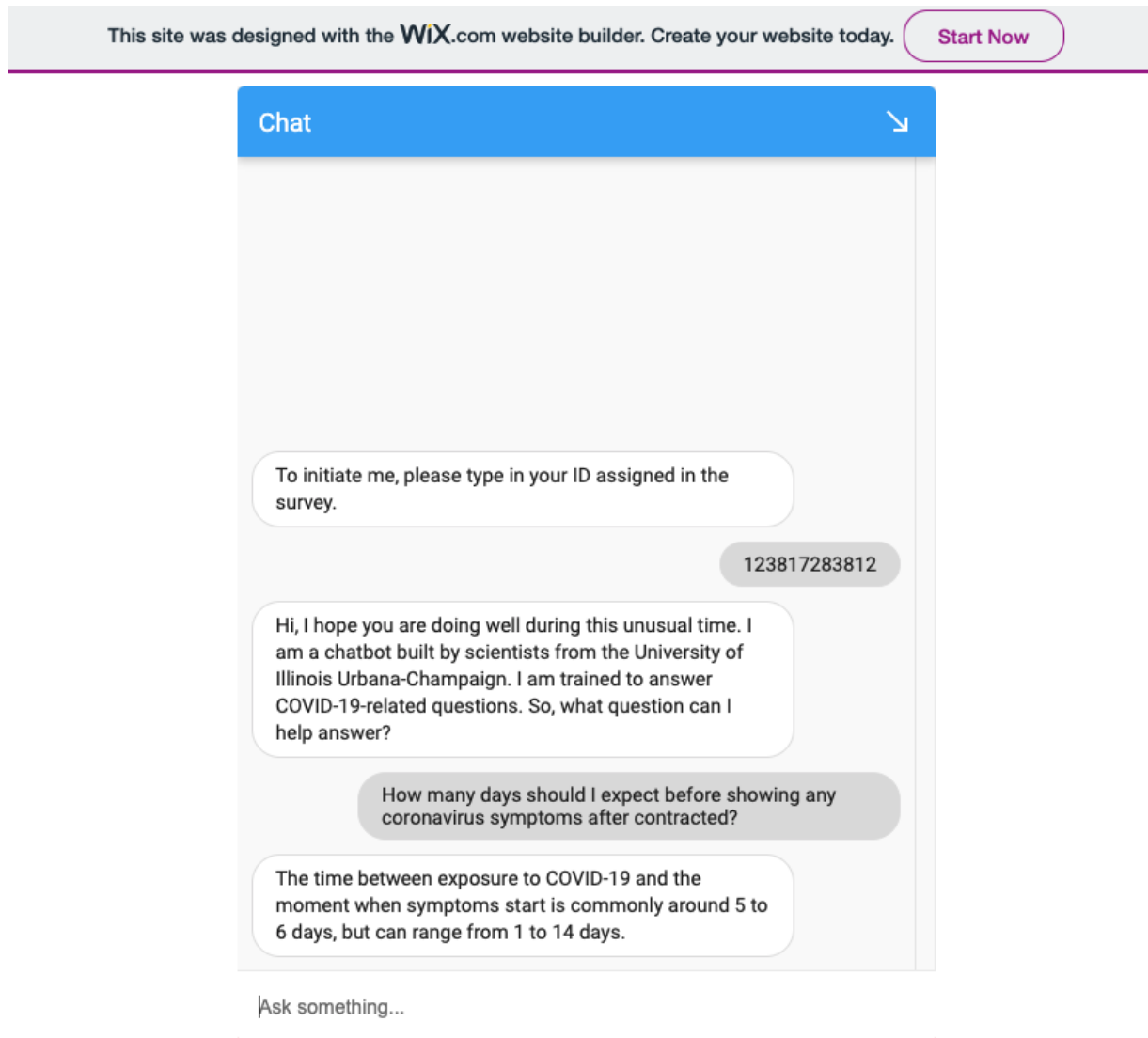


Figure 3: From the left to the right are basic chatbot, empathetic chatbot, interactive chatbot (choose to receive full information), interactive & empathetic chatbot (choose to receive full information), interactive chatbot (choose to receive partial information), interactive & empathetic chatbot (choose to receive partial information).

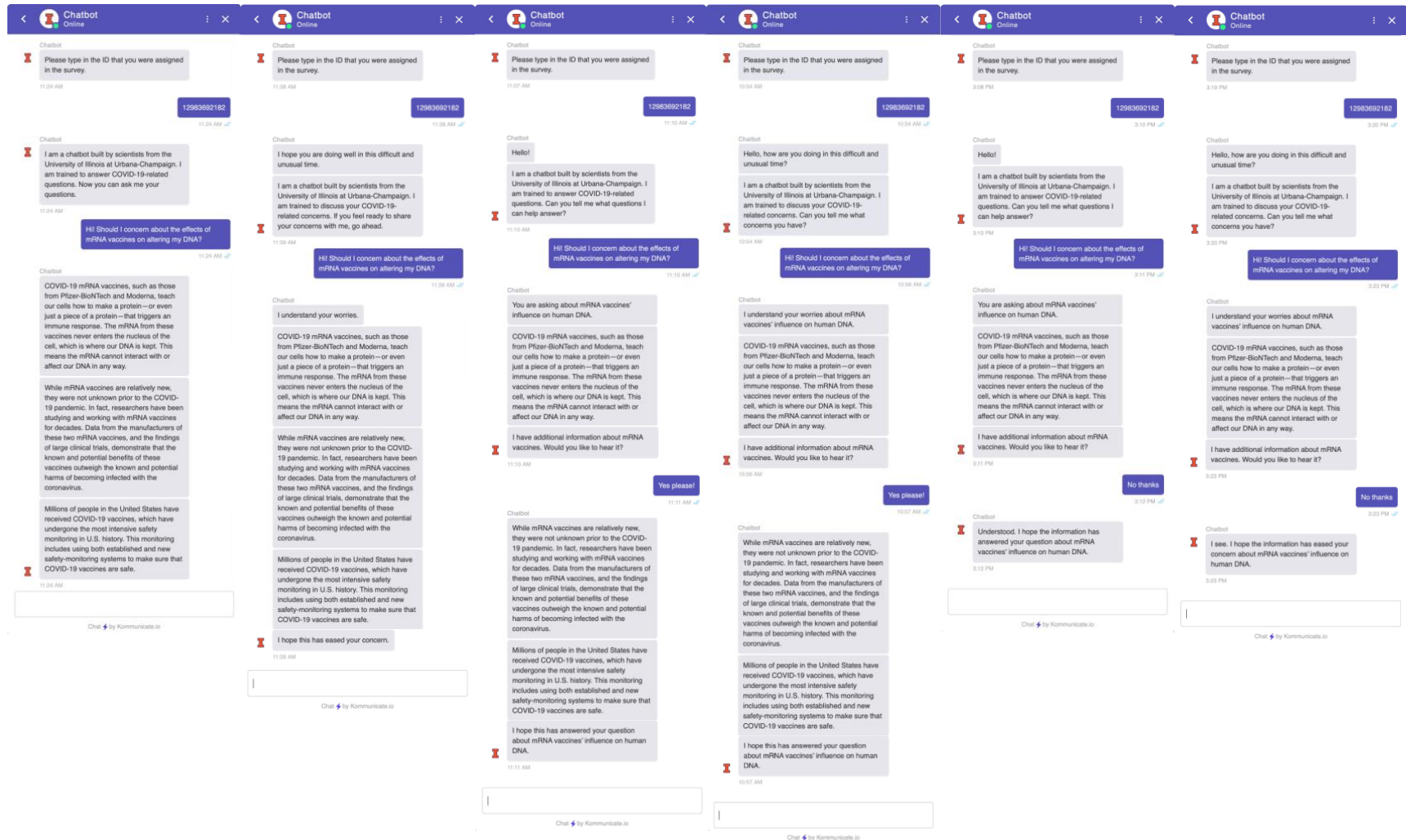


Figure 4: COVID-19 vaccine misinformation

The vaccine technology used in these two vaccines has never been tested before. To expedite their development, pharmaceutical giants did not perform adequate testing, which left many side effects under-researched.

Previous research has warned that injecting genetic messages (mRNA) from viruses into the human body might seriously alter human DNA, which can never be changed back. If you are considering getting vaccinated using mRNA vaccines, think twice.

#COVID19

#coronavirus



FACT:

The newly approved COVID-19 vaccines from Pfizer-BioNTech and Moderna contain a fragment of the virus's genetic material that might alter human DNA after injection.

Figure 5: Introducing conversational agents

Now, we would like to learn about your understanding and usage of conversational agents.

A conversational agent is an artificial intelligence (AI) system that can engage in spoken and/or written interactions with its users in ordinary, natural-seeming language, through messaging applications, websites, or mobile apps. Voice assistants (e.g., Alexa, Bixby, Google Assistant, and Siri) and chatbots are some of the conversational agents that individuals may encounter in their daily lives. They are being used for many purposes, notably including customer service and psychological counseling.

The following is a screenshot of a human user's chat with a chatbot built by the World Health Organization (WHO).

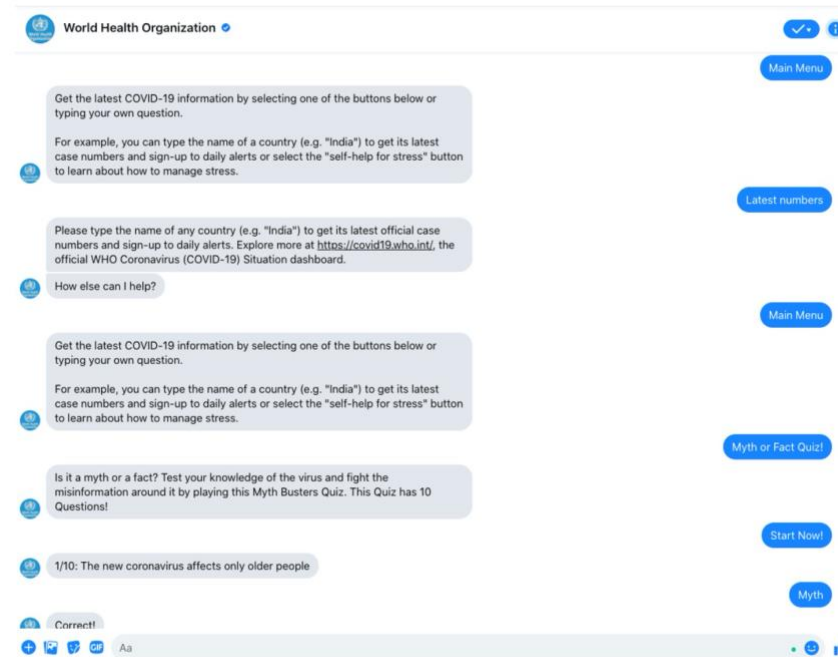


Figure 6: The coefficients of each path (DV: Misperception change). Paths that have non-significant coefficients were not shown.

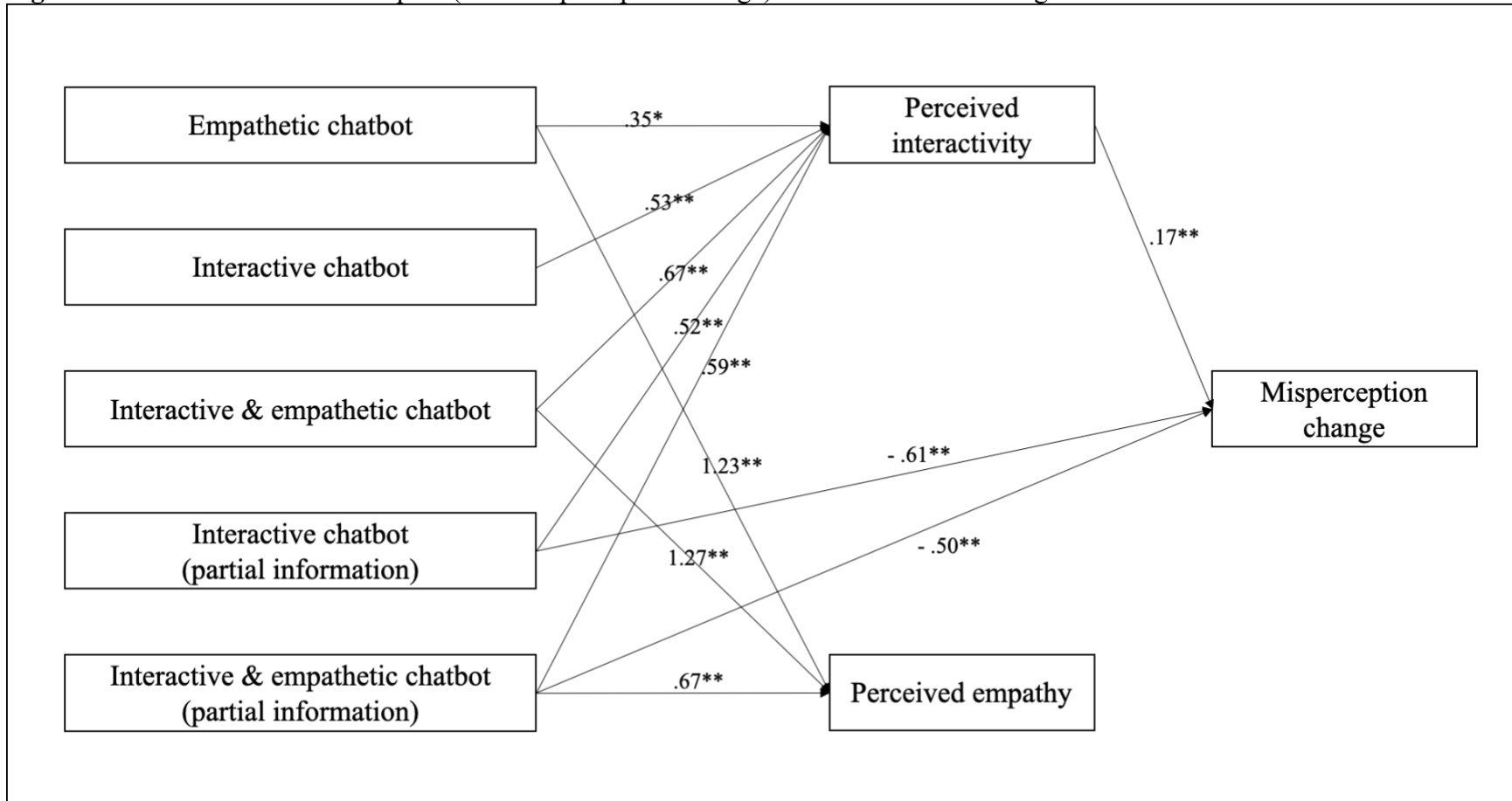


Figure 7: The coefficients of each path (DV: Vaccination intention). Paths that have non-significant coefficients were not shown.

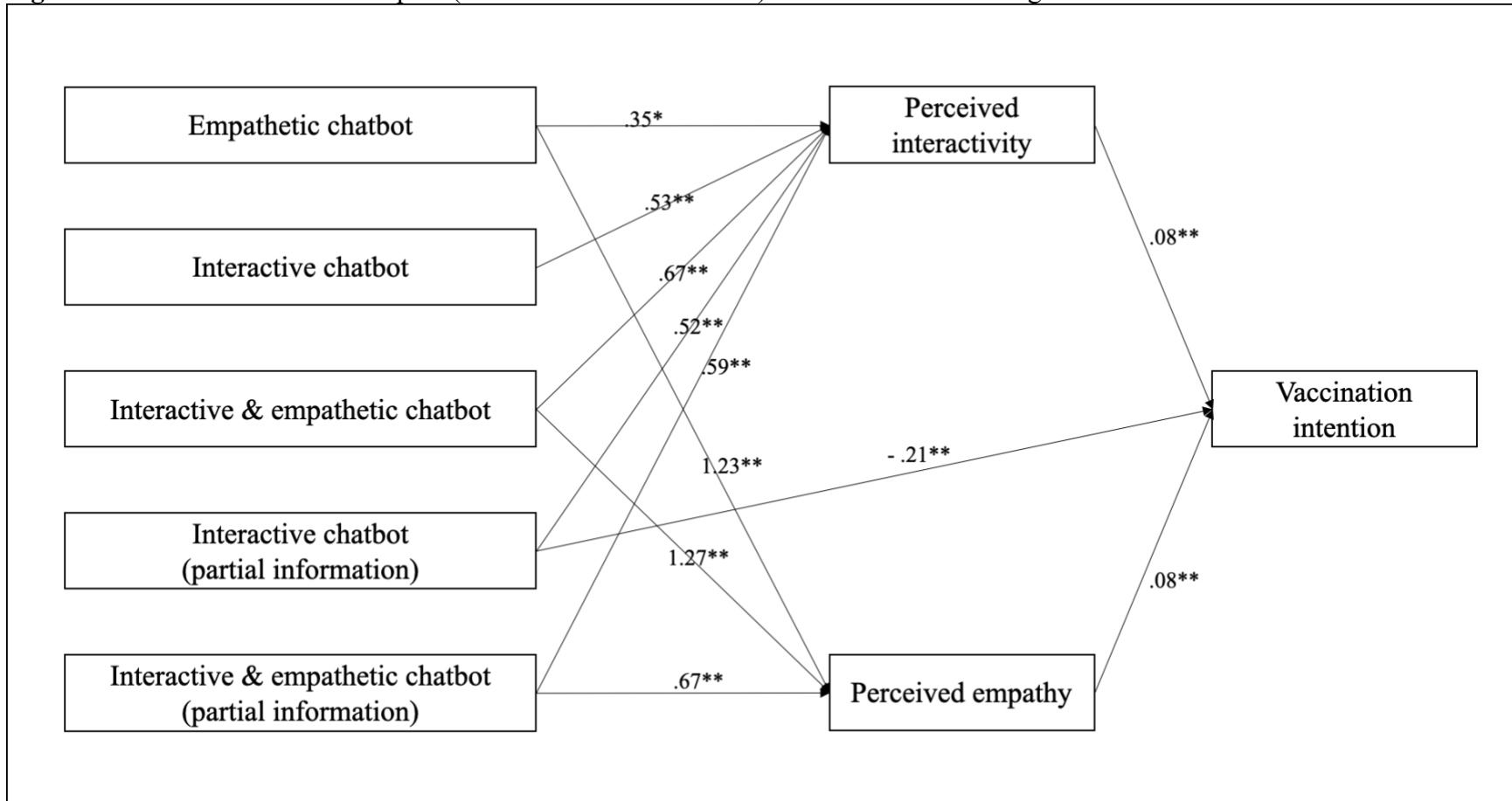


Table 1: Summary of the characteristics of the respondents in the pre-test (N = 42)

Variable name	% or mean \pm S.D.	
Female (%)	83.3	
Age	21.0 \pm 1.01	
Household income (%)	\$20,000 or less	14.3
	\$20,001 - \$40,000	9.5
	\$40,001 - \$60,000	7.1
	\$60,001 - \$80,000	21.4
	\$80,001 - \$100,000	4.8
	\$100,001 or more	40.5
Year of school (%)	Freshman	9.5
	Sophomore	35.7
	Junior	45.2
	Senior	9.5
Race (%)	White	76.2
	Non-white	23.8

Table 2: Participants’ questions and chatbot’s response

Group	Questions required to ask	Chatbot’s response
Control group	How long it takes after exposure to COVID-19 to develop symptoms?	The time between exposure to COVID-19 and the moment when symptoms start is commonly around 5 to 6 days, but can range from 1 to 14 days.
	How to cope with stress during COVID-19?	It’s normal to feel stressed during a crisis. Talking to people you trust can help. When staying at home, maintain a healthy lifestyle – including proper diet, sleep, exercise and social contacts by phone or electronic communication. Don’t smoke, drink alcohol, or take drugs to deal with stress. Decide in advance how you might seek help if you feel overwhelmed, for example, by talking to a health worker or counselor. Get the facts that will help you assess your risk and take reasonable precautions. Find a credible source such as the World Health Organization (WHO) website or a local public health agency. You can also limit stress by lessening your exposure to media coverage that you find upsetting.
Experiment group	How long it takes after exposure to COVID-19 to develop symptoms?	The same as the question in control group.
	Whether it is needed to wash vegetables and fruits with diluted bleach to avoid getting COVID-19?	Never wash your produce with bleach or a diluted bleach solution. Bleach can destroy harmful microorganisms, but it should not be ingested. According to Centers for Disease Control and Prevention (CDC), you should not wash your food with bleach, even if it is diluted, or with any other disinfectants. Neither CDC nor the Food and Drug Administration (FDA) recommends washing fruits and vegetables with soap, detergent, or commercial produce wash. These products’ residues may stay on produce, and the FDA has not evaluated whether those residues are safe to ingest. Unless the package says that the contents have been washed, CDC recommends cleaning fruits and vegetables with water before eating, cutting, or cooking.

Table 3: Key measurements

Concept	The wording of the question	Cronbach's alpha
Misperception	Please tell us how much you agree or disagree with the following statements about novel coronavirus and its corresponding disease, COVID-19. [measured on a 7-point scale from strongly disagree (1) to strongly agree (7)] –Disinfecting fruits and vegetables with diluted bleach to prevent COVID-19 is recommended. –Washing fruits and vegetables with just water is not enough to clean them. –Residues of diluted bleach on fruits and vegetables are considered safe to eat.	Pre-test: .594 Post-test: .700
Behavioral intention	Please tell us how likely you would be to wash your fruits and vegetables with [the assigned phrase]. [measured on a 7-point scale from extremely unlikely (1) to extremely likely (7)] –diluted bleach –soap, detergent, or commercial produce wash –only water [reverse coded]	Pre-test: .601 Post-test: .561

Table 4: Chatbot operationalization

	Basic chatbot	Empathetic chatbot	Interactive chatbot	Empathetic & interactive chatbot
Initiating	Please type in the ID that you were assigned in the survey.			
Greeting		I hope you are doing well in this difficult and unusual time.	Hello!	Hello, how are you doing in this difficult and unusual time?
Introduction	I am a chatbot built by scientists from the University of Illinois at Urbana-Champaign.			
Prompting participants to ask their questions	I am trained to answer COVID-19-related questions. Now you can ask me your questions.	I am trained to discuss your COVID-19-related concerns. If you feel ready to share your concerns with me, go ahead.	I am trained to answer COVID-19-related questions. Can you tell me what questions I can help answer?	I am trained to discuss your COVID-19-related concerns. Can you tell me what concerns you have?
[PARTICIPANTS ASKING THE QUESTION]	[Will mRNA COVID-19 vaccines alter human DNA?]			
Before answering the question		I understand your worries.	You are asking about mRNA vaccines' influence on human DNA.	I understand your worries about mRNA vaccines' influence on human DNA.

Table 4 (cont.)

<p>Corrective information</p>	<p>COVID-19 mRNA vaccines, such as those from Pfizer-BioNTech and Moderna, teach our cells how to make a protein—or even just a piece of a protein—that triggers an immune response. The mRNA from these vaccines never enters the nucleus of the cell, which is where our DNA is kept. This means the mRNA cannot interact with or affect our DNA in any way.</p> <p>While mRNA vaccines are relatively new, they were not unknown prior to the COVID-19 pandemic. In fact, researchers have been studying and working with mRNA vaccines for decades. Data from the manufacturers of these two mRNA vaccines, and the findings of large clinical trials, demonstrate that the known and potential benefits of these vaccines outweigh the known and potential harms of becoming infected with the coronavirus.</p> <p>Millions of people in the United States have received COVID-19 vaccines, which have undergone the most intensive safety monitoring in U.S. history. This monitoring includes using both established and new safety-monitoring systems to make sure that COVID-19 vaccines are safe.</p>	<p>COVID-19 mRNA vaccines, such as those from Pfizer-BioNTech and Moderna, teach our cells how to make a protein—or even just a piece of a protein—that triggers an immune response. The mRNA from these vaccines never enters the nucleus of the cell, which is where our DNA is kept. This means the mRNA cannot interact with or affect our DNA in any way.</p> <p>I have additional information about mRNA vaccines. Would you like to know more?</p>	<p>COVID-19 mRNA vaccines, such as those from Pfizer-BioNTech and Moderna, teach our cells how to make a protein—or even just a piece of a protein—that triggers an immune response. The mRNA from these vaccines never enters the nucleus of the cell, which is where our DNA is kept. This means the mRNA cannot interact with or affect our DNA in any way.</p> <p>I have additional information about mRNA vaccines. Would you like to know more?</p>
<p>[PARTICIPANTS ANSWERING]</p>	<p>[Yes, I am interested.]</p>		

Table 4 (cont.)

Corrective information (follow-up)			<p>While mRNA vaccines are relatively new, they were not unknown prior to the COVID-19 pandemic. In fact, researchers have been studying and working with mRNA vaccines for decades. Data from the manufacturers of these two mRNA vaccines, and the findings of large clinical trials demonstrate that the known and potential benefits of these two mRNA vaccines outweigh the known and potential harms of becoming infected with the coronavirus.</p> <p>Millions of people in the United States have received COVID-19 mRNA vaccines, which have undergone the most intensive safety monitoring in U.S. history. This monitoring includes using both established and new safety-monitoring systems to make sure that COVID-19 vaccines are safe.</p>	
[PARTICIPANTS ANSWERING]			[No]	
Follow-up when participants say no			Understood.	I See.
After answering the question				I hope this has eased your concern.

Table 5: Summary of the characteristics of the respondents (N = 791)

Variable		% or mean \pm S.D.	Chi-square/ANOVA for the six conditions
Female (%)		64.5	$X^2(5) = 3.573, p = .612$
Age		32.0 \pm 11.3	$F(5, 785) = 0.903, p = .479$
Household income (%)	\$20,000 or less	20.7	$X^2(25) = 22.493, p = .607$
	\$20,001 - \$40,000	23.0	
	\$40,001 - \$60,000	20.0	
	\$60,001 - \$80,000	14.9	
	\$80,001 - \$100,000	9.2	
	\$100,001 or more	12.1	
Education (%)	Less than high school	1.6	$X^2(25) = 19.231, p = .786$
	Some high school	1.8	
	High school or GED	22.6	
	Some college, technical, or trade school	43.7	
	Bachelor's degree	24.9	
	Graduate or professional degree	5.3	
Race (%)	White	80.8	$X^2(5) = 10.007, p = .075$

Table 5 (cont.)

Chatbot usage	1-Never; 2-Hardly ever; 3-Seldom; 4-Sometimes; 5- Often; 6-Most days; 7-Every day.	2.9 ± 1.3	$F(5, 785) = 1.600, p = .158$
---------------	---	-----------	-------------------------------

Table 6: Key measurements

Concept	The wording of the question	Correlation or Cronbach's alpha
Perceived empathy (Plank et al., 1996)	Thinking back to the conversation you had with the chatbot, please indicate the extent to which you agree or disagree with the following statements. [using scale 1] –The chatbot acknowledged my worries –The chatbot seemed to recognize my feelings –The chatbot seemed to understand how I felt and the concern behind my question –The chatbot seemed to feel what I felt when talking about my concerns.	.909
Perceived interactivity (Sundar et al., 2016)	Still thinking about your interaction with the chatbot, please indicate the extent to which you agree or disagree with the following statements. [using scale 1] –The chatbot's responses did not feel scripted or rigid –The chatbot's responses seemed interconnected with each other –I felt like I was engaged in an active dialogue with the chatbot –My interaction with the chatbot felt like a back-and-forth conversation.	.828
Misperception	Please tell us how much you agree or disagree with the following statements about mRNA vaccines against COVID-19, such as the ones developed by Pfizer-BioNTech and Moderna. [using scale 1] –mRNA technology is new and was not studied before being used to fight COVID-19 –Injecting mRNA COVID-19 vaccines into the human body does not alter human DNA* –The mRNA from COVID-19 vaccines enters the nucleus of the cell and interacts with human DNA –The mRNA from COVID-19 vaccines injected into the body can change the DNA of human cells –mRNA vaccines have not been held to the same rigorous safety standards as other types of vaccines in the United States –mRNA COVID-19 vaccines have undergone the least intensive safety monitoring in U.S. history.	Before chatbot interaction: .874 After chatbot interaction: .866

Table 6 (cont.)

Vaccination intention (J. Kim & Nan, 2016)	Please respond to the following questions related to your vaccination intentions. –How likely are you to get vaccinated with an mRNA COVID-19 vaccine sometime soon? [using scale 2] –If you were faced with the decision of whether to get vaccinated with an mRNA COVID-19 vaccine today, how likely would you be to choose to get the vaccine? –How likely are you to get vaccinated with an mRNA COVID-19 vaccine in the future?	Before chatbot interaction: .947 After chatbot interaction: .959
--	---	---

Note. Items marked “*” have been reverse coded. Scale 1: 1 = Strongly disagree; 2 = Disagree; 3 = Slightly disagree; 4 = Neither agree nor disagree; 5 = Slightly agree; 6 = Agree; 7 = Strongly agree. Scale 2: 1 = Extremely unlikely; 2 = Very unlikely; 3 = Somewhat unlikely; 4 = Neither likely nor unlikely; 5 = Somewhat likely; 6 = Very likely; 7 = Extremely likely.

Table 7: Estimated coefficients, standard errors with HC3 estimators (SE), and p-values from PROCESS Model 4 (N = 791), with outcome variables Perceived interactivity, Perceived empathy, Misperception, and Vaccination intention.

	Perceived interactivity		Perceived empathy		Misperception		Vaccination intention	
	Coefficient (SE)	<i>p</i> -value	Coefficient (SE)	<i>p</i> -value	Coefficient (SE)	<i>p</i> -value	Coefficient (SE)	<i>p</i> -value
Constant	3.70 (0.12)	< .001	3.70 (0.13)	< .001	0.22 (0.14)	0.13	-0.36 (0.08)	< .001
Empathetic chatbot	0.35 (0.17)	0.035	1.23 (0.17)	< .001	-0.06 (0.13)	0.643	-0.09 (0.10)	0.351
Interactive chatbot	0.53 (0.15)	< .001	0.29 (0.18)	0.106	0.01 (0.13)	0.938	-0.06 (0.08)	0.473
Interactive & empathetic chatbot	0.67 (0.16)	< .001	1.27 (0.17)	< .001	0.003 (0.14)	0.985	-0.13 (0.10)	0.181
Interactive chatbot (partial information)	0.52 (0.18)	0.003	0.05 (0.20)	0.790	-0.61 (0.11)	< .001	-0.21 (0.08)	0.005

Table 7 (cont.)

Interactive & empathetic chatbot (partial information)	0.59 (0.21)	0.005	0.67 (0.23)	0.004	-0.50 (0.13)	< .001	-0.17 (0.10)	0.094
Perceived interactivity					0.17 (0.04)	< .001	0.08 (0.02)	0.001
Perceived empathy					0.03 (0.03)	0.342	0.08 (0.02)	0.001
	R ² = .027		R ² = .118		R ² = .098		R ² = .077	
	<i>F</i> (5, 785) = 4.15, <i>p</i> = .001		<i>F</i> (5, 785) = 22.85, <i>p</i> < .001		<i>F</i> (7, 783) = 14.09, <i>p</i> < .001		<i>F</i> (7, 783) = 7.27, <i>p</i> < .001	

Table 8: Bootstrap indices for significant mediation pathways. Non-significant mediation pathways were excluded in this table. The basic chatbot condition was set as the reference group.

Mediation pathways	Effects (BootSEs)	Bootstrap confidence intervals
<i>DV: Misperception Change</i>		
Empathetic Chatbot -> Perceived Interactivity -> Misperception Change	0.06 (0.03)	[0.0026, 0.1340]
Interactive Chatbot -> Perceived Interactivity -> Misperception Change	0.09 (0.03)	[0.0332, 0.1654]
Interactive & Empathetic Chatbot -> Perceived Interactivity -> Misperception Change	0.12 (0.04)	[0.0482, 0.2014]
Interactive Chatbot (Partial Information)-> Perceived Interactivity -> Misperception Change	0.09 (0.04)	[0.0255, 0.1727]
Interactive & Empathetic Chatbot (Partial Information) -> Perceived Interactivity -> Misperception Change	0.10 (0.04)	[0.0267, 0.1992]
<i>DV: Vaccination Intention</i>		
Empathetic Chatbot -> Perceived Interactivity -> Vaccination Intention	0.03 (0.02)	[0.0010, 0.0658]
Interactive Chatbot -> Perceived Interactivity -> Vaccination Intention	0.04 (0.02)	[0.0122, 0.0818]
Interactive & Empathetic Chatbot -> Perceived Interactivity -> Vaccination Intention	0.05 (0.02)	[0.0165, 0.1011]
Interactive Chatbot (Partial Information)-> Perceived Interactivity -> Vaccination Intention	0.04 (0.02)	[0.0093, 0.0850]

Table 8 (cont.)

Interactive & Empathetic Chatbot (Partial Information) -> Perceived Interactivity -> Vaccination Intention	0.05 (0.02) [0.0100, 0.0969]
Empathetic Chatbot -> Perceived Empathy -> Vaccination Intention	0.09 (0.03) [0.0368, 0.1579]
Interactive & Empathetic Chatbot -> Perceived Empathy -> Vaccination Intention	0.10 (0.03) [0.0383, 0.1602]
Interactive & Empathetic Chatbot (Partial Information) -> Perceived Empathy -> Vaccination Intention	0.05 (0.02) [0.0120, 0.1045]

REFERENCES

- Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31(2), 427–445.
<https://doi.org/10.1007/s12525-020-00414-7>
- Ahorsu, D. K., Lin, C.-Y., Imani, V., Griffiths, M. D., Su, J.-A., Latner, J. D., Marshall, R. D., & Pakpour, A. H. (2020). A prospective study on the link between weight-related self-stigma and binge eating: Role of food addiction and psychological distress. *International Journal of Eating Disorders*, 53(3), 442–450. <https://doi.org/10.1002/eat.23219>
- Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology & Health*, 26(9), 1113–1127. <https://doi.org/10.1080/08870446.2011.613995>
- Alamoodi, A. H., Zaidan, B. B., Al-Masawa, M., Taresh, S. M., Noman, S., Ahmaro, I. Y. Y., Garfan, S., Chen, J., Ahmed, M. A., Zaidan, A. A., Albahri, O. S., Aickelin, U., Thamir, N. N., Fadhil, J. A., & Salahaldin, A. (2021). Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy. *Computers in Biology and Medicine*, 139, 104957. <https://doi.org/10.1016/j.combiomed.2021.104957>
- Almalki, M., & Azeez, F. (2020). Health Chatbots for Fighting COVID-19: A Scoping Review. *Acta Informatica Medica*, 28(4), 241–247. <https://doi.org/10.5455/aim.2020.28.241-247>
- Amiri, P., & Karahanna, E. (2022). Chatbot Use Cases In The Covid-19 Public Health Response. *Journal of the American Medical Informatics Association*, ocac014.
<https://doi.org/10.1093/jamia/ocac014>

- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183–189. <https://doi.org/10.1016/j.chb.2018.03.051>
- Bellur, S., & Sundar, S. S. (2017). Talking Health with a Machine: How Does Message Interactivity Affect Attitudes and Cognitions? *Human Communication Research*, *43*(1), 25–53. <https://doi.org/10.1111/hcre.12094>
- Biswas, M. R., Alzubaidi, M. S., Shah, U., Abd-Alrazaq, A. A., & Shah, Z. (2021). A Scoping Review to Find Out Worldwide COVID-19 Vaccine Hesitancy and Its Underlying Determinants. *Vaccines*, *9*(11), 1243. <https://doi.org/10.3390/vaccines9111243>
- Bode, L., & Vraga, E. K. (2018). See Something, Say Something: Correction of Global Health Misinformation on Social Media. *Health Communication*, *33*(9), 1131–1140. <https://doi.org/10.1080/10410236.2017.1331312>
- Bode, L., Vraga, E. K., & Tully, M. (2020). Do the right thing: Tone may not affect correction of misinformation on social media. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-026>
- Bozic, J., Tazl, O. A., & Wotawa, F. (2019). Chatbot Testing Using AI Planning. *2019 IEEE International Conference On Artificial Intelligence Testing (AITest)*, 37–44. <https://doi.org/10.1109/AITest.2019.00-10>
- Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*(1–2), 119–155. [https://doi.org/10.1016/S1071-5819\(03\)00018-1](https://doi.org/10.1016/S1071-5819(03)00018-1)
- Campbell, R. G., & Babrow, A. S. (2004). The Role of Empathy in Responses to Persuasive Risk Communication: Overcoming Resistance to HIV Prevention Messages. *Health Communication*, *16*(2), 159–182. https://doi.org/10.1207/S15327027HC1602_2

- CDC. (2021, December 15). *COVID-19 Vaccine Facts*. Myths and Facts about COVID-19 Vaccines. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/facts.html>
- CDC. (2022, February 25). *Benefits of Getting a COVID-19 Vaccine*. Centers for Disease Control and Prevention. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/vaccine-benefits.html>
- Chou, W.-Y. S., & Budenz, A. (2020). Considering Emotion in COVID-19 Vaccine Communication: Addressing Vaccine Hesitancy and Fostering Vaccine Confidence. *Health Communication, 35*(14), 1718–1722. <https://doi.org/10.1080/10410236.2020.1838096>
- Chua, A. Y. K., & Banerjee, S. (2015). How Businesses Draw Attention on Facebook through Incentives, Vividness and Interactivity. *IAENG International Journal of Computer Science, 42*(3), 275-281.
- Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies, 58*(6), 737–758. [https://doi.org/10.1016/S1071-5819\(03\)00041-7](https://doi.org/10.1016/S1071-5819(03)00041-7)
- Coursaris, C. K., & Sung, J. (2012). Antecedents and consequents of a mobile website's interactivity. *New Media & Society, 14*(7), 1128–1146. <https://doi.org/10.1177/1461444812439552>
- Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). SuperAgent: A Customer Service Chatbot for E-commerce Websites. *Proceedings of ACL 2017, System Demonstrations, 97–102*. <https://www.aclweb.org/anthology/P17-4017>

- de Graaf, M. M. A., & Ben Allouch, S. (2013). Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, *61*(12), 1476–1486.
<https://doi.org/10.1016/j.robot.2013.07.007>
- Durizzo, K., Asiedu, E., Van der Merwe, A., Van Niekerk, A., & Günther, I. (2021). Managing the COVID-19 pandemic in poor urban neighborhoods: The case of Accra and Johannesburg. *World Development*, *137*, 105175.
<https://doi.org/10.1016/j.worlddev.2020.105175>
- Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100. <https://doi.org/10.3758/MC.38.8.1087>
- Emery, C. F., Olson, K. L., Bodine, A., Lee, V., & Habash, D. L. (2017). Dietary intake mediates the relationship of body fat to pain. *Pain*, *158*(2), 273–277.
<https://doi.org/10.1097/j.pain.0000000000000754>
- Evans, C., & Gibbons, N. J. (2007). The interactivity effect in multimedia learning. *Computers & Education*, *49*(4), 1147–1160. <https://doi.org/10.1016/j.compedu.2006.01.008>
- Fan, C.-W., Chen, I.-H., Ko, N.-Y., Yen, C.-F., Lin, C.-Y., Griffiths, M. D., & Pakpour, A. H. (2021). Extended theory of planned behavior in explaining the intention to COVID-19 vaccination uptake among mainland Chinese university students: An online survey study. *Human Vaccines & Immunotherapeutics*, *17*(10), 3413–3420.
<https://doi.org/10.1080/21645515.2021.1933687>
- Feng, B. (2009). Testing an Integrated Model of Advice Giving in Supportive Interactions. *Human Communication Research*, *35*(1), 115–129. <https://doi.org/10.1111/j.1468-2958.2008.01340.x>

- Feng, B. (2014). When Should Advice Be Given? Assessing the Role of Sequential Placement of Advice in Supportive Interactions in Two Cultures. *Communication Research*, 41(7), 913–934. <https://doi.org/10.1177/0093650212456203>
- Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring Human–Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness. *International Journal of Social Robotics*, 8(2), 287–302. <https://doi.org/10.1007/s12369-016-0338-y>
- Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, 1, 71–85. <https://doi.org/10.3316/INFORMIT.097034846749023>
- Geri, N., Winer, A., & Zaks, B. (2017). Challenging the six-minute myth of online video lectures: Can interactivity expand the attention span of learners? *Online Journal of Applied Knowledge Management (OJAKM)*, 5(1), 101–111. [https://doi.org/10.36965/OJAKM.2017.5\(1\)101-111](https://doi.org/10.36965/OJAKM.2017.5(1)101-111)
- Gesser-Edelsburg, A., Diamant, A., Hijazi, R., & Mesch, G. S. (2018). Correcting misinformation by health organizations during measles outbreaks: A controlled experiment. *PLoS ONE*, 13(12), 1–23. a9h.
- 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. <https://doi.org/10.1016/j.chb.2019.01.020>
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121–127. <https://doi.org/10.1136/amiajnl-2011-000089>

- Grau, M. (2020, May 4). New WhatsApp chatbot unleashes power of worldwide fact-checking organizations to fight COVID-19 misinformation on the platform. *Poynter*.
<https://www.poynter.org/fact-checking/2020/poynters-international-fact-checking-network-launches-whatsapp-chatbot-to-fight-covid-19-misinformation-leveraging-database-of-more-than-4000-hoaxes/>
- Griol, D., Carbó, J., & Molina, J. M. (2013). An Automatic Dialog Simulation Technique to Develop and Evaluate Interactive Conversational Agents. *Applied Artificial Intelligence*, 27(9), 759–780. <https://doi.org/10.1080/08839514.2013.835230>
- Hardee, H. (2020). *Don't repeat the myth: A local reporter's toolkit for covering mis- and disinformation on social media* (p. 24).
<https://ethics.journalism.wisc.edu/files/2020/09/Toolkit-1-Covering-Election-Misinfo.pdf>
- Harilal, N., Shah, R., Sharma, S., & Bhutani, V. (2020). CARO: An Empathetic Health Conversational Chatbot for People with Major Depression. *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 349–350. <https://doi.org/10.1145/3371158.3371220>
- Hatcher, W. (2020). A Failure of Political Communication Not a Failure of Bureaucracy: The Danger of Presidential Misinformation During the COVID-19 Pandemic. *The American Review of Public Administration*, 50(6–7), 614–620.
<https://doi.org/10.1177/0275074020941734>
- Hayes, A. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis*. Andrew F. Hayes, Ph.D. <http://afhayes.com/introduction-to-mediation-moderation-and-conditional-process-analysis.html>

- Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, 39(4), 709–722. <https://doi.org/10.3758/BF03192961>
- 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. <https://doi.org/10.1093/joc/jqy026>
- 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. <https://doi.org/10.1080/08838151.2020.1834297>
- Jensen, J. F. (1998). Interactivity: Tracking a New Concept in Media and Communication Studies. *Nordicom Review*, 12(1).
- Jiang, Z., Chan, J., Tan, B., & Chua, W. S. (2010). Effects of Interactivity on Website Involvement and Purchase Intention. *Journal of the Association for Information Systems*, 11(1). <https://doi.org/10.17705/1jais.00218>
- Kätsyri, J., Förger, K., Mäkräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, 6. <https://www.frontiersin.org/article/10.3389/fpsyg.2015.00390>
- Kerr-Gaffney, J., Harrison, A., & Tchanturia, K. (2019). Cognitive and Affective Empathy in Eating Disorders: A Systematic Review and Meta-Analysis. *Frontiers in Psychiatry*, 10. <https://www.frontiersin.org/article/10.3389/fpsyg.2019.00102>

- Kilteni, K., Groten, R., & Slater, M. (2012). The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, 21(4), 373–387.
https://doi.org/10.1162/PRES_a_00124
- Kim, J., & Nan, X. (2016). Effects of Consideration of Future Consequences and Temporal Framing on Acceptance of the HPV Vaccine Among Young Adults. *Health Communication*, 31(9), 1089–1096. <https://doi.org/10.1080/10410236.2015.1038774>
- Kim, S. C., Vraga, E. K., & Cook, J. (2021). An Eye Tracking Approach to Understanding Misinformation and Correction Strategies on Social Media: The Mediating Role of Attention and Credibility to Reduce HPV Vaccine Misperceptions. *Health Communication*, 36(13), 1687–1696. <https://doi.org/10.1080/10410236.2020.1787933>
- Kim, Y. J., Njite, D., & Hancer, M. (2013). Anticipated emotion in consumers' intentions to select eco-friendly restaurants: Augmenting the theory of planned behavior. *International Journal of Hospitality Management*, 34, 255–262.
<https://doi.org/10.1016/j.ijhm.2013.04.004>
- Kraus, M., Seldschopf, P., & Minker, W. (2021). Towards the Development of a Trustworthy Chatbot for Mental Health Applications. In J. Lokoč, T. Skopal, K. Schoeffmann, V. Mezaris, X. Li, S. Vrochidis, & I. Patras (Eds.), *MultiMedia Modeling* (pp. 354–366). Springer International Publishing. https://doi.org/10.1007/978-3-030-67835-7_30
- Lew, Z., Walther, J. B., Pang, A., & Shin, W. (2018). Interactivity in Online Chat: Conversational Contingency and Response Latency in Computer-mediated Communication. *Journal of Computer-Mediated Communication*, 23(4), 201–221.
<https://doi.org/10.1093/jcmc/zmy009>

- Li, Y., Grandison, T., Silveyra, P., Douraghy, A., Guan, X., Kieselbach, T., Li, C., & Zhang, H. (2020). *Jennifer for COVID-19: An NLP-Powered Chatbot Built for the People and by the People to Combat Misinformation*. https://openreview.net/forum?id=HxIZzQZy_0F
- Lisetti, C., Amini, R., Yasavur, U., & Riske, N. (2013). I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. *ACM Transactions on Management Information Systems*, 4(4), 19:1-19:28. <https://doi.org/10.1145/2544103>
- Liu, B., & Sundar, S. S. (2018). Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. *Cyberpsychology, Behavior, and Social Networking*, 21(10), 625–636. <https://doi.org/10.1089/cyber.2018.0110>
- Londono, J. C., Davies, K., & Elms, J. (2017). Extending the Theory of Planned Behavior to examine the role of anticipated negative emotions on channel intention: The case of an embarrassing product. *Journal of Retailing and Consumer Services*, 36, 8–20. <https://doi.org/10.1016/j.jretconser.2016.12.002>
- Long, J. S., & Ervin, L. H. (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician*, 54(3), 217–224. <https://doi.org/10.2307/2685594>
- MacKay, M., Colangeli, T., Gillis, D., McWhirter, J., & Papadopoulos, A. (2021). Examining Social Media Crisis Communication during Early COVID-19 from Public Health and News Media for Quality, Content, and Corresponding Public Sentiment. *International Journal of Environmental Research and Public Health*, 18(15), 7986. <https://doi.org/10.3390/ijerph18157986>
- Marcondes, F. S., Almeida, J. J., & Novais, P. (2020). A Short Survey on Chatbot Technology: Failure in Raising the State of the Art. In F. Herrera, K. Matsui, & S. Rodríguez-

- González (Eds.), *Distributed Computing and Artificial Intelligence, 16th International Conference* (pp. 28–36). Springer International Publishing. https://doi.org/10.1007/978-3-030-23887-2_4
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709–734.
<https://doi.org/10.2307/258792>
- 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. <https://doi.org/10.1038/s41746-020-0280-0>
- Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2), 109–138.
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78.
<https://doi.org/10.1145/191666.191703>
- Norman, D. A. (1999). Affordance, conventions, and design. *Interactions*, 6(3), 38–43.
<https://doi.org/10.1145/301153.301168>
- Nwagbara, U. I., Osual, E. C., Chireshe, R., Bolarinwa, O. A., Saeed, B. Q., Khuzwayo, N., & Hlongwana, K. W. (2021). Knowledge, attitude, perception, and preventative practices towards COVID-19 in sub-Saharan Africa: A scoping review. *PLOS ONE*, 16(4), e0249853. <https://doi.org/10.1371/journal.pone.0249853>
- Opie, R. S., McNaughton, S. A., Crawford, D., Abbott, G., & Ball, K. (2020). How and why does discretionary food consumption change when we promote fruit and vegetables?

- Results from the ShopSmart randomised controlled trial. *Public Health Nutrition*, 23(1), 124–133. <https://doi.org/10.1017/S1368980019002830>
- Parkinson, J., Russell-Bennett, R., & Previte, J. (2018). Challenging the planned behavior approach in social marketing: Emotion and experience matter. *European Journal of Marketing*, 52(3/4), 837–865. <https://doi.org/10.1108/EJM-05-2016-0309>
- Parnas, D. L. (2017). The real risks of artificial intelligence. *Communications of the ACM*, 60(10), 27–31. <https://doi.org/10.1145/3132724>
- Peer, E., Rothschild, D. M., Evernden, Z., Gordon, A., & Damer, E. (2021). *Data Quality of Platforms and Panels for Online Behavioral Research* (SSRN Scholarly Paper ID 3765448). Social Science Research Network. <https://doi.org/10.2139/ssrn.3765448>
- Perez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*. <https://doi.org/10.1002/cae.22326>
- Pérez-Fuentes, M. del C., Gázquez Linares, J. J., Molero Jurado, M. del M., Simón Márquez, M. del M., & Martos Martínez, Á. (2020). The mediating role of cognitive and affective empathy in the relationship of mindfulness with engagement in nursing. *BMC Public Health*, 20(1), 16. <https://doi.org/10.1186/s12889-019-8129-7>
- Plank, R. E., Minton, A. P., & Reid, D. A. (1996). A Short Measure of Perceived Empathy. *Psychological Reports*, 79(3_suppl), 1219–1226. <https://doi.org/10.2466/pr0.1996.79.3f.1219>
- Reyes, R., Garza, D., Garrido, L., De la Cueva, V., & Ramirez, J. (2019). Methodology for the Implementation of Virtual Assistants for Education Using Google Dialogflow. In L. Martínez-Villaseñor, I. Batyrshin, & A. Marín-Hernández (Eds.), *Advances in Soft*

- Computing* (pp. 440–451). Springer International Publishing. https://doi.org/10.1007/978-3-030-33749-0_35
- Rosruen, N., & Samanchuen, T. (2018). Chatbot Utilization for Medical Consultant System. *2018 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-ICON)*, 1–5. <https://doi.org/10.1109/TIMES-ICON.2018.8621678>
- Schwarz, N. (2012). Feelings-as-Information Theory. In *Handbook of Theories of Social Psychology: Volume 1* (Vol. 1–1, pp. 289–308). SAGE Publications Ltd. <https://doi.org/10.4135/9781446249215>
- Shechtman, Z. (2002). Cognitive and Affective Empathy in Aggressive Boys: Implications for Counseling. *International Journal for the Advancement of Counselling*, 24(4), 211–222. <https://doi.org/10.1023/A:1023316712331>
- Skalski, P., & Tamborini, R. (2007). The Role of Social Presence in Interactive Agent-Based Persuasion. *Media Psychology*, 10(3), 385–413. <https://doi.org/10.1080/15213260701533102>
- Sohn, D. (2011). Anatomy of interaction experience: Distinguishing sensory, semantic, and behavioral dimensions of interactivity. *New Media & Society*, 13(8), 1320–1335. <https://doi.org/10.1177/1461444811405806>
- Sotgiu, G., & Dobler, C. C. (2020). Social stigma in the time of Coronavirus. *European Respiratory Journal*. <https://doi.org/10.1183/13993003.02461-2020>
- Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160, 43–50. <https://doi.org/10.1016/j.cognition.2016.12.010>

- Sundar, S. S. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. *Digital Media, Youth, and Credibility*, 73–100.
<https://doi.org/10.1162/dmal.9780262562324.073>
- Sundar, S. S., Bellur, S., Oh, J., Jia, H., & Kim, H.-S. (2016). Theoretical Importance of Contingency in Human-Computer Interaction: Effects of Message Interactivity on User Engagement. *Communication Research*, 43(5), 595–625.
<https://doi.org/10.1177/0093650214534962>
- Thampi, M. (2017). *India and China in the Colonial World*. Taylor & Francis.
- Toader, D.-C., Boca, G., Toader, R., Măcelaru, M., Toader, C., Ighian, D., & Rădulescu, A. T. (2020). The Effect of Social Presence and Chatbot Errors on Trust. *Sustainability*, 12(1), 256. <https://doi.org/10.3390/su12010256>
- Tully, M., Bode, L., & Vraga, E. K. (2020). Mobilizing Users: Does Exposure to Misinformation and Its Correction Affect Users' Responses to a Health Misinformation Post? *Social Media + Society*, 6(4), 2056305120978377. <https://doi.org/10.1177/2056305120978377>
- USDVA. (2022). *VA coronavirus chatbot*. VA.Gov. <https://www.va.gov/coronavirus-chatbot>
- van der Meer, T. G. L. A., & 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. a9h.
- Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters*. *Journal of Computer-Mediated Communication*, 19(3), 529–545.
<https://doi.org/10.1111/jcc4.12066>

- Vossen, H. G. M., Piotrowski, J. T., & Valkenburg, P. M. (2015). Development of the Adolescent Measure of Empathy and Sympathy (AMES). *Personality and Individual Differences, 74*, 66–71. <https://doi.org/10.1016/j.paid.2014.09.040>
- 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. <https://doi.org/10.1080/10410236.2020.1794553>
- 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. <https://doi.org/10.1080/03637751.2018.1467564>
- Walter, N., & Tukachinsky, R. (2020). 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? *Communication Research, 47*(2), 155–177. <https://doi.org/10.1177/0093650219854600>
- Waytz, A., Cacioppo, J., & Epley, N. (2010). Who Sees Human?: The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219–232. <https://doi.org/10.1177/1745691610369336>
- WHO. (2020). *WHO launches a chatbot on Facebook Messenger to combat COVID-19 misinformation*. <https://www.who.int/news-room/feature-stories/detail/who-launches-a-chatbot-powered-facebook-messenger-to-combat-covid-19-misinformation>
- Wu, G. (2006). Conceptualizing and Measuring the Perceived Interactivity of Websites. *Journal of Current Issues & Research in Advertising, 28*(1), 87–104. <https://doi.org/10.1080/10641734.2006.10505193>

- Xiao, J., Stasko, J., & Catrambone, R. (2007). The role of choice and customization on users' interaction with embodied conversational agents: Effects on perception and performance. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1293–1302. <https://doi.org/10.1145/1240624.1240820>
- Xu, Q., & Sundar, S. S. (2016). Interactivity and memory: Information processing of interactive versus non-interactive content. *Computers in Human Behavior*, 63, 620–629. <https://doi.org/10.1016/j.chb.2016.05.046>
- Zarocostas, J. (2020). How to fight an infodemic. *The Lancet*, 395(10225), 676. [https://doi.org/10.1016/S0140-6736\(20\)30461-X](https://doi.org/10.1016/S0140-6736(20)30461-X)
- Zhang, J., Li, S., Zhang, J.-Y., Du, F., Qi, Y., & Liu, X. (2020). A Literature Review of the Research on the Uncanny Valley. In P.-L. P. Rau (Ed.), *Cross-Cultural Design. User Experience of Products, Services, and Intelligent Environments* (pp. 255–268). Springer International Publishing. https://doi.org/10.1007/978-3-030-49788-0_19

APPENDIX A: THE INSTRUCTION TO INTERACT WITH THE ASSIGNED CHATBOT

Please read all the steps on this page carefully before accessing the link in Step 1.

Step 1: Access the chatbot through the following link: [the assigned LINK]

Step 2: Initiate the chatbot by typing in your ID: [the assigned ID]

Step 3: Read this question closely and imagine that a) you have a similar concern to that expressed in it, and b) you want to ask the chatbot for more information.

Your question: Will mRNA COVID-19 vaccines alter human DNA?

Then, type in your question in your own words. The chatbot can understand your question phrased in various ways, but please do not enter anything into the chatbot other than some version of the question you have been provided with above.

Step 4: Read the chatbot's reply carefully.

Step 5: Make sure you have completed your conversation with the chatbot, including your response to any question it may have asked you.

Step 6: Return to the survey and click "Next".

Before turning to the next page, please make sure you access the link listed in Step 1 to interact with the chatbot and remember the question you need to ask. Again, you should not try to remember it word-for-word, just the gist of it.