Brief Communications

User reactions to COVID-19 screening chatbots from reputable providers

Alan R. Dennis[1*], Antino Kim[1], Mohammad Rahimi[2], Sezgin Ayabakan[2]
[1]Kelley School of Business, Indiana University, Bloomington, IN
[2]Fox School of Business, Temple University, Philadelphia, PA

[*]Correspondence to Alan Dennis, ardennis@indiana.edu
Kelley School of Business, Indiana University, Bloomington IN 47405
812-855-2691

Keywords: Public health, information technology, chatbot, health screening

Word Count: 1921

ABSTRACT

Objective: The objective was to understand how people respond to COVID-19 screening chatbots.

Materials and Methods: We conducted an online experiment with 371 participants who viewed a COVID-19 screening session between a hotline agent (chatbot or human) and a user with mild or severe symptoms.

Results: The primary factor driving user response to screening hotlines (human or chatbot) is perceptions of the agent's ability. When ability is the same, users view chatbots no differently or *more positively* than human agents. The primary factor driving perceptions of ability is the user's trust in the hotline provider, with a slight negative bias against chatbots' ability. Asians perceived higher ability and benevolence than Whites.

Conclusion: Ensuring that COVID-19 screening chatbots provide high quality service is critical, but not sufficient for widespread adoption. The key is to emphasize the chatbot's ability and assure users that it delivers the *same* quality as human agents.

INTRODUCTION

Many people are seeking information in response to the COVID-19 pandemic [1]. Individuals with various symptoms and conditions are looking for guidance on whether to seek medical attention for COVID-19. Providing accurate, timely information is crucial to help those with—as well as those without—COVID-19 make good decisions. The sudden unprecedented demand for information is overwhelming resources [2, 3]. One solution is the deployment and use of technologies such as chatbots [3, 4].

Chatbots have the potential to relieve the pressure on contact centers [3, 5]. Chatbots are software applications that conduct an online conversation in natural language via typed text or voice commands (e.g., Siri) [6]. Chatbots are scalable, so they can meet an unexpected surge in demand when there is a shortage of qualified human agents [7]. Chatbots can provide round-the-clock service at a low operational cost [7]. They are consistent in quality in that they always provide the same results in response to the same inputs, and are easily retrained in the face of rapidly changing information [8]. Chatbots are also non-judgmental; they make no moral judgments about the information provided by the user, so users may be more willing to disclose socially undesirable information [9].

As chatbots increase in quality, their use is expanding. For example, chatbots are already widely deployed in customer service applications to guide users through knowledge bases or well-structured processes (e.g., technical and customer supports) [9]. Chatbots integrate directly into existing web, phone, social media and message channels, and can be launched in many different languages [10].

Chatbots are increasingly being deployed in healthcare [11, 12]. The COVID-19 pandemic has spurred even greater deployment, many for screening of potential patients [3, 13]. COVID-19 screening is an ideal application for chatbots because it is a well-structured process that involves asking patients a series of clearly-defined questions and determining a risk score [9, 14]. Chatbots can help call centers triage patients and advise them on the most appropriate actions to take, which may be to do nothing because the patient does not present symptoms that warrant immediate medical care [14].

Despite all the potential benefits, like any other technology-enabled services, chatbots will help *only if people use them and follow their advice* [11, 15]. In this paper, we examine whether people will use high-quality chatbots provided by reputable organizations. We control for chatbot quality by examining a chatbot that provides the exact same service as a human agent. COVID-19 screening is based on a very specific set of criteria, so a well-designed chatbot can perform at close to a trained human level [16].

Trust is an important factor that influences the use of chatbots [11], as well as patient compliance [17, 18]. Users will be reluctant to use chatbots if they do not trust them [11]. Trust in humans is influenced by three primary factors [19] that also have parallels for trust in technology [20]. The first is *ability*: the agent—human or chatbot—must be competent within the range of actions required of it [19]. The agent must have the knowledge and skills needed to make a correct diagnosis. Second, *integrity*: the agent must do what it says it will do [19]. For example, if the agent says the user's information is private and will not be disclosed, the information must truly be private. In the era where data breaches are common [21], do users believe that technology has integrity? Finally, *benevolence*: the agent must have the patient's best interests in mind, and not be guided by ulterior motives, such as increasing profits [19].

The underlying trust factors of ability, integrity, and benevolence play important roles in the use of technology, and technology providing recommendations in particular [22-24]. Ability and integrity are typically more important for instrumental outcomes associated with transactions (e.g., purchasing) because users are most concerned with whether the technology will work as intended to complete the transaction [22-24]. Affect and other perceptual outcomes (e.g., satisfaction) are often influenced more by benevolence as these are based more on relationship aspects of technology use [22-24]. Accordingly, we examine ability, integrity, and benevolence as potential factors to drive trust in chatbots and, subsequently, influence patients' intentions to use chatbots and comply with their recommendations.

METHOD

We conducted a 2×2 between-subjects—two agent types (human vs chatbot) by two patient severity levels (mild vs severe)--online experiment where subjects were randomly assigned to view a video vignette of COVID-19 screening hotline session between an agent and a patient. The online setting is appropriate as screening services can be provided via various online channels [10, 13]. Vignettes have been commonly used to study human behavior [25], technology use [26], and trust [27] because they provide excellent experimental control [28]. Research shows that reading or watching a vignette triggers the same attitudes as actually engaging in the behaviors shown in the vignette [25]; meta-analyses have shown no significant differences in conclusions between vignette studies and studies of actual behavior, although effect sizes in vignette-based studies tend to be slightly lower [25, 26].

In April, 2020, we recruited 402 participants from Amazon Mechanical Turk following usual protocols to ensure data quality [29]. Participants were paid $2.00. Thirty subjects failed one or more of the six attention checks and one did not report gender, and were removed, leaving 371 participants for analysis. About half were female (188), 83% were White, 8% Asian, 6% Black and 3% other (individuals selecting multiple ethnicities and individuals selecting "other"). The median age was 40 with most participants aged 25-64 (1%: 18-24; 26%: 25-34; 34%: 35-44; 19%: 45-54; 15%: 55-64; 5%: 65 or more). There were no significant differences in gender, age or race across the four conditions. The Supplementary Materials provide the detailed demographics by condition.

Participants watched a 2½ minute video vignette of a fictitious text chat between an agent at a COVID-19 screening hotline and a user with possible COVID-19 symptoms. We designed two vignettes in which the users either reported mild or severe symptoms. We developed our vignettes based on our experiences using four COVID-19 chatbots [13] and the screening questions recommended by the CDC. Participants were informed that the video was either a human agent or a chatbot (randomly assigned), but the videos were the same between the two conditions to control for quality differences between human and chatbot. Thus, the study compares a chatbot with capabilities identical in quality to those of a human agent. Participants were informed that the hotline was provided by the Centers for Disease Control and Prevention (CDC) and were informed of the deception at the end of the study. Thus, any differences between the chatbot and human agent are due to human bias because participants saw the exact same vignette in both conditions.

We used established measures of ability, integrity, benevolence, trust, and the control factors of disposition to trust, and personal innovativeness with information technology. We adapted prior measures for satisfaction, persuasiveness, likelihood of use and likelihood of following up on the diagnosis of the agent. All measures used 1-7 scales and all scales proved reliable (Cronbach alpha > .80). All demographic items were categorical variables. More details on the items and reliabilities are provided in the Supplementary Materials. The experimental materials were pilot tested with 100 undergraduate students at the first author's university prior to the study.

RESULTS

The first part of our analysis shows that participants perceived the chatbot to have significantly less ability, integrity and benevolence (see Table 1). Severity of symptoms influenced the perceptions of ability and integrity, but not benevolence. The effect sizes for the models as a whole (R^2) were what Cohen [30] calls medium or small to medium. The individual effect sizes of the chatbot (partial eta^2) for ability and integrity were between what Cohen [30] terms small (.01) and medium (.06), while the effect size for benevolence was medium. The primary factor influencing perceptions of ability was trust in the provider (i.e., the CDC), with the type of agent (human or chatbot) being a secondary factor. For integrity, both the trust in the provider and the type of agent were primary factors. For benevolence, the primary factor was the type of agent, with trust secondary. We also controlled for gender, age, and ethnicity. Gender had no significant effect but compared to Whites, individuals of Asian ethnicity perceived the agent to have significantly higher ability and benevolence. Age was significant for benevolence but there was no pattern to its effects.

In the second part of our analysis, we examined five outcomes: (i) persuasiveness, (ii) satisfaction, (iii) likelihood of following the agent's advice, (iv) trust, and (v) likelihood of use (see Table 2), after controlling for the effects of ability, integrity and benevolence. The effect sizes for the models as a whole (R^2) were large. The dominant factor across all five outcomes was perceived ability (very large effect sizes), with chatbot a secondary factor having a medium-sized *positive* effect on persuasiveness, and small to medium *positive* effects on satisfaction, likelihood of following the agent's advice, and likelihood of use. Lastly, severity of the condition did not directly affect the outcomes nor moderate the relationship between chatbot and outcomes. The control variables (gender, age, and ethnicity) had no significant effects on the outcome variables.

DISCUSSION

Simply put, the results show that the primary factor driving patient response to COVID-19 screening hotlines (human or chatbot) is users' perceptions of the agent's ability. A secondary factor for persuasiveness, satisfaction, likelihood of following the agent's advice, and likelihood of use was the type of agent, with participants reporting they viewed chatbots *more positively* than human agents, which is good news for healthcare organizations struggling to meet user demand for screening services. This positive response may be because users feel more comfortable disclosing information to a chatbot, especially socially undesirable information, because a chatbot makes no judgment [9]. The CDC, the World Health Organization (WHO), UNICEF and other health organizations caution that the COVID-19 outbreak has provoked social stigma and discriminatory behaviors against people of certain ethnic backgrounds, as well as those perceived to have been in contact with the virus [31, 32]. This is truly an unfortunate situation, and perhaps chatbots can assist those who are hesitant to seek help because of the stigma.

The primary factor driving perceptions of ability was the user's trust in the provider of the screening hotline. Our results show a slight negative bias against chatbots' ability, perhaps due to recent press reports [13]. Therefore, proactively informing users of the chatbot's ability is important; users need to understand that chatbots use the same up-to-date knowledge base and follow the same set of screening protocols as human agents.

CONCLUSION

Developing a high-quality COVID-19 screening chatbot—as qualified as a trained human agent—will help alleviate the increased load on COVID-19 contact centers staffed by human agents. When chatbots are perceived to provide the same service quality as human agents, users are more likely to see them as persuasive, be more satisfied, and be more likely to use them. A user's tech-savviness (PIIT) has only a small effect, so these results apply to both those with deep technology experience and those with little.

Yet, therein lies the rub: There is a gap between how users perceive chatbots' and human agents' abilities. Therefore, to offset users' biases [33], a necessary component in deploying chatbots for COVID-19 screening is a strong messaging campaign that emphasizes the chatbot's ability. Because trust in the provider strongly influences perceptions of ability, building on the organization's reputation may also prove useful.

Table 1. Results for ability, integrity and benevolence showing beta coefficients

	Ability	Integrity	Benevolence
Chatbot	−0.399***	−0.435***	−0.616***
Severe Symptoms	0.136*	0.297**	0.329
Chatbot × Severe Symptoms	0.103	0.003	−0.260
Higher Risk Participant	0.030	0.013	0.013
Disposition to Trust	0.162***	0.218***	0.202**
Personal Innovativeness with IT (PIIT)	0.108*	0.126*	0.164*
Trust in CDC	0.331***	0.221***	0.217**
Female	0.109	0.001	0.136
Age	Included	Included	Included*
Ethnicity	Included*	Included	Included*
Constant	6.125***	4.511***	4.650***
R^2	0.269	0.216	0.193
Adjusted R^2	0.234	0.178	0.154
F	5.363	5.101	8.434
Effect Sizes (Partial *eta*^{2})			
Chatbot	0.042	0.045	0.088
Severe Symptoms	0.012	0.021	0.007
Chatbot x Severe Symptoms	0.001	0.000	0.003
Higher Risk	0.001	0.000	0.000
Disposition to Trust	0.031	0.037	0.023
PIIT	0.016	0.015	0.017
Trust in CDC	0.120	0.040	0.027
Female	0.004	0.000	0.003
Age	0.030	0.024	0.039
Ethnicity	0.023	0.005	0.026

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Table 2. Results for outcomes showing beta coefficients

	Persuasive	Satisfaction	Follow Advice	Trust	Likely to Use
Chatbot	0.272***	0.112***	0.035*	0.022	0.270**
Severe Symptoms	0.097	0.044	−0.143	0.088	0.004
Chatbot × Severe Symptoms	0.014	0.069	0.268	0.026	0.039
Higher Risk Participant	−0.024	−0.024	−0.039	0.001	0.000
Disposition to Trust	0.015	0.035	0.016	-0.006	0.051
Personal Innovativeness with IT (PIIT)	0.028	0.021	0.038	0.043	0.115*
Trust in CDC	−0.001	0.030	0.238***	0.071*	0.087
Female	−0.058	0.005	0.048	-0.125	−0.031
Age	Included	Included	Included	Included	Included
Ethnicity	Included	Included	Included	Included	Included
Ability	0.583***	0.603***	0.634***	0.612***	0.786***
Integrity	0.105**	0.049	−0.006	0.350***	0.070
Benevolence	0.084*	0.005	0.105	0.072	0.300***
Constant	5.605***	5.82***	6.883***	6.191***	5.949***
R^2	0.671	0.766	0.553	0.741	0.594
Adjusted R^2	0.653	0.752	0.527	0.726	0.571
F	35.759	57.167	21.633	50.140	25.601
Effect Sizes (Partial *eta*^{2})					
Chatbot	0.065	0.034	0.011	0.001	0.022
Severe Symptoms	0.010	0.010	0.000	0.007	0.000
Chatbot x Severe Symptoms	0.000	0.002	0.007	0.000	0.000
Higher Risk Participant	0.002	0.004	0.002	0.000	0.000
Disposition to Trust	0.001	0.007	0.000	0.000	0.002
PIIT	0.003	0.003	0.002	0.005	0.014
Trust in CDC	0.000	0.005	0.068	0.011	0.007
Female	0.003	0.000	0.001	0.010	0.000
Age	0.010	0.010	0.022	0.008	0.016
Ethnicity	0.007	0.004	0.001	0.009	0.004
Ability	0.410	0.576	0.266	0.373	0.277
Integrity	0.016	0.007	0.000	0.126	0.002
Benevolence	0.011	0.000	0.008	0.006	0.042

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Funding Statement: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing Interests Statement: The authors have no competing interests to declare.

Protection of Human Subjects Statement: The study was reviewed by the Indiana University Institutional Review Board as protocol #2003099481 and was determined to be an exempt study.

Contributorship Statement: A.R.D. and K.A. conceptualized the research. S.A., A.R.D., A.K., and M.R. designed the methodology. A.K. collected and curated the data. S.A., A.R.D. and A.K., analyzed the data, S.A., A.R.D., A.K., and M.R. wrote the paper.

REFERENCES

1. Drees J. Led by COVID-19 surge, virtual visits will surpass 1B in 2020: report. Becker's Hospital Review 2020(5/18/2020)
2. Brodwin E. Surge in patients overwhelms telehealth services amid coronavirus pandemic. STAT 2020(5/18/2020)
3. Judson TJ, Odisho AY, Young JJ, et al. Case Report: Implementation of a Digital Chatbot to Screen Health System Employees during the COVID-19 Pandemic. Journal of the American Medical Informatics Association 2020
4. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. npj Digital Medicine 2020;**3**(1):1-4
5. Hao K. The pandemic is emptying call centers. AI chatbots are swooping in. MIT Technology Review 2020(5/18/2020)
6. Mauldin M. ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition. Proceedings of the Eleventh National Conference on Artificial Intelligence: AAAI Press, 1994:16-21.
7. Mittal A, Agrawal A, Chouksey A, Shriwas R, Agrawal S. A Comparative Study of Chatbots and Humans. International Journal of Advanced Research in Computer and Communications Engineering 2016;**5**(3):1055-57
8. AbuShawar B, Atwell E. Usefulness, localizability, humanness, and language-benefit: additional evaluation criteria for natural language dialogue systems. Int Journal of Speech Technology 2016;**19**:373–83
9. Følstad A, Skjuve M. Chatbots for customer service: user experience and motivation. 1st International Conference on Conversational User Interfaces New York: Association for Computing Machinery,, 2019:1–9.
10. A New Chatbot for Customer Service on Social Media. CHI'17; 2017; Denver, CO.
11. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digit Health 2019;**5**:1-12 doi: 10.1177/2055207619871808[published Online First: Epub Date]|.
12. Chatbots in Healthcare: Status Quo, Application Scenarios for Physicians and Paitents, and Future Directions. European Conference on Information Systems; 2020; Marakesh.
13. Ross C. I asked eight chatbots whether I had Covid-19. The answers ranged from 'low' risk to 'start home isolation'. STAT 2020

14. Ghosh S, Bhatia S, Bhatia A. Quro: Facilitating User Symptom Check Using a Personalised Chatbot-Oriented Dialogue System. Studies in health technology and informatics 2018;**252**:51-56
15. Gefen D, Karahanna E, Straub DW. Trust and TAM in Online Shopping: An Integrated Model. MIS Quarterly 2003;**27**(1):51-90
16. Adiwardana D, Luong M-T, So DR, et al. Towards a human-like open-domain chatbot. arXiv preprint arXiv:2001.09977 2020
17. Lu X, Zhang R, Wu W, Shang X, Liu M. Relationship Between Internet Health Information and Patient Compliance Based on Trust: Empirical Study. Journal of Medical Internet Resesearch 2018;**20**(8):e253
18. Kerse N, Buetow S, Mainous AG, Young G, Coster G, Arroll B. Physician-Patient Relationship and Medication Compliance: A Primary Care Investigation. The Annals of Family Medicin2 2004;**2**(455-461)
19. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Academy of management review 1995;**20**(3):709-34
20. Lankton NK, McKnight DH, Tripp J. Technology, Humanness, and Trust: Rethinking Trust in Technology. Journal of the Association for Information Systems 2015;**16**(10):880-918
21. Liu V, Musen MA, Chou T. Data breaches of protected health information in the United States. Journal of the American Medical Association 2015;**313**(4):1471-73
22. Benbasat I, Wang W. Trust in and adoption of online recommendation agents. Journal of the association for information systems 2005;**6**(3):4
23. Wang W, Qiu L, Kim D, Benbasat I. Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. Decision Support Systems 2016;**86**:48-60 doi: https://doi.org/10.1016/j.dss.2016.03.007[published Online First: Epub Date]|.
24. Xu J, Cenfetelli RT, Aquino K. Do different kinds of trust matter? An examination of the three trusting beliefs on satisfaction and purchase behavior in the buyer–seller context. The Journal of Strategic Information Systems 2016;**25**(1):15-31 doi: https://doi.org/10.1016/j.jsis.2015.10.004[published Online First: Epub Date]|.
25. Shaw JC, Wild RE, Colquitt JA. To justify or excuse? A meta-analysis of the effects of explanations. Journal of Applied Psychology 2003;**88**(3):444-58
26. Cram AW, D'Arcy J, Proudfoot JG. Seeing the Forest and the Trees: A Meta-Analysis of the Antecedents to Information Security Polict Compliance. MIS Quarterly 2019;**43**(2):525-54
27. Elsbach K, Elofson G. How the packaging of decision explanations affects perceptions of trustworthiness. Academy of Management Journal 2000;**43**(1):80-89
28. Aguinis H, Bradley KJ. Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. Organizational Research Methods 2014;**17**(4):351-71
29. Steelman ZR, Hammer BI, Limayem M. Data Collection in the Digital Age: Innovative Alternatives to Student Samples. MIS Quarterly 2014;**38**(2):355-78
30. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
31. CDC. Reducing Stigma. Secondary Reducing Stigma 2020. https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/reducing-stigma.html.
32. UNICEF. Social stigma associated with the coronavirus disease (COVID-19). Secondary Social stigma associated with the coronavirus disease (COVID-19) 2020.

https://www.unicef.org/documents/social-stigma-associated-coronavirus-disease-covid-19.
33. Luo X, Tong S, Fang Z, Qu Z. Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. Marketing Science 2019;**38**(6):937-47

Supplementary Text

I. Constructs

What follows is the list of questions and statements used in the survey for each construct. Reliability information is provided at the end of this document.

Age
What is your age?
- 1: 18-24
- 2: 25-34
- 3: 35-44
- 4: 45-54
- 5: 55-64
- 6: 65-74
- 7: 75-84
- 8: 85 or older

Gender
To which gender identity do you most identify?

Race
Please specify your ethnicity.
- 1: White, Caucasian, Middle Eastern, North African
- 2: Black or African American
- 3: American Indian or Alaska Native
- 4: Asian/ Asian American
- 5: Native Hawaiian or Pacific Islander
- 6: Native American, Inuit or Aluet
- 7: Other

Note: For our analysis, we have grouped the races as:
- *1: Caucasian*
- *2: Black*
- *4: Asian*
- *Else: Other*

Disposition to Trust (Source: [1-5])
Please indicate whether you agree or disagree with the following statements. (1-7 scale)
- Most people are honest in describing their experience and abilities.
- Most people tell the truth about the limits of their knowledge.
- Most people can be counted on to do what they say they will do.
- Most people answer personal questions honestly.
- Most people are competent in terms of their work.

PIIT (Personal Innovativeness in Information Technology) (Source: [6])
Please indicate whether you agree or disagree with the following statements. (1-7 scale)

- If I heard about a new information technology, I would look for ways to experiment with it.
- Among my peers, I am usually the first to try out new information technologies.
- In general, I am hesitant to try out new information technologies.
- I like to experiment with new information technologies.

Trust in the CDC

- How trustworthy is the CDC (Centers for Disease Control and Prevention)? (1-7 scale)

Risk Group
(Source: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html)
Note: Participants were classified as "higher risk" if they satisfied any of the following conditions.

- Do you live in a nursing home or long-term care facility?
- Do you have any of the following medical conditions?
  - Immunocompromised or an organ transplant recipient
  - Cancer
  - Diabetes
  - Lung disease (COPD, emphysema, pulmonary fibrosis, bronchitis, asthma requiring daily inhaler use)
  - Am pregnant
  - None of the above
- Do you have a history of any of the following?
  - Major surgery or fracture within the past two months
  - Heart, kidney, or liver disease
  - History of blood clot in legs or lungs
  - IV drug use
  - None of the above

Satisfaction (Source: [7-11]) (1-7 scale)
Please indicate whether you agree or disagree with the following statements.

- Are you satisfied with Robin?
- Did Robin meet the user's needs?
- Are you satisfied with the accuracy of the service provided by Robin?
- Was Robin clear to understand?
- Was Robin easy to talk to?
- Did Robin perform satisfactorily?
- Was the service provided by Robin accurate?

Trust – Ability (Source: [12]) (1-7 scale)
Please indicate whether you agree or disagree with the following statements.

- Robin seemed to be successful in the activities he/she undertook.
- I felt very confident about Robin's skills.
- Robin was well qualified.

Trust – Benevolence (Source: [12]) (1-7 scale)
Please indicate whether you agree or disagree with the following statements.

- Robin was concerned about what was important to the user.
- Robin cared about the user's feelings.
- Robin was benevolent.

Trust – Integrity (Source: [12]) (1-7 scale)
Please indicate whether you agree or disagree with the following statements.

- If Robin said he/she was going to do something, he/she did it.
- I like Robin's work values.
- Robin showed integrity.

Trust in Agent (Source: [12]) (1-7 scale)
Please indicate whether you agree or disagree with the following statements.

- Robin can be trusted to make sensible decisions
- Overall, Robin is very trustworthy.
- I trust Robin.
- I can rely on Robin.
- I lack confidence in Robin.

Persuasiveness (Source: [13]) (1-7 scale)
To what extent do you agree or disagree that the information Robin provided was

- Believable
- Convincing
- Important to the user
- Helpful in making the user feel confident about what to do
- Successful in making the user want to follow instructions
- Agreeable

Follow Advice (Source: [14]) (1-7 scale)

- How likely would you be to take advice from Robin?
- How likely would you be to take advice from Robin again in the future (if a similar situation took place)?
- How likely would you be to follow up/carry through with the next steps proposed by Robin?
- How soon would you be willing to carry through with the next steps proposed by Robin?

Likelihood of Use / Intention to Use (Source: [15]) (1-7 scale)
Please indicate whether you agree or disagree with the following statements.

- If I was faced with a similar situation, I would interact with Robin.
- If I was faced with a similar decision in the future, I would contact Robin.
- If a similar need arises in the future, I would feel comfortable contacting Robin to meet my needs.
- If I had problems like this, I would contact Robin.

II. Attention Check Questions

What follows is the list of questions and statements used as attention checks in the survey.

A.
Question 3.1_6: Choose "Somewhat agree" if you're paying attention.
Question 6.3: Choose "Like a little" if you are paying attention.
Question 11.1_8: Choose Somewhat agree if you're paying attention

B.
The following questions were used as attention checks right after the videos, depending on the treatment that the subject had received.

(Treatment: Chatbot Robin – Severe)
Question 7.2: True or false: Robin is a **chatbot** at Covid-19 Screening Hotline.
Question 7.3: True or false: This video **has been speeded up** for this experiment.
Question 7.4: True or false: In the video, the user works at a pharmacy.

(Treatment: Chatbot Robin – Mild)
Question 8.2: True or false: Robin is a **chatbot** at Covid-19 Screening Hotline.
Question 8.3: True or false: This video **has been speeded up** for this experiment.
Question 8.4: True or false: In the video, the user works at a pharmacy.

(Treatment: Human Robin - Severe)
Question 9.2: True or false: Robin is a **qualified staff member** at Covid-19 Screening Hotline.
Question 9.3: True or false: This video **has been speeded up** for this experiment.
Question 9.4: True or false: In the video, the user works at a pharmacy.

(Treatment: Human Robin - Mild)
Question 10.2: True or false: Robin is a **qualified staff member** at Covid-19 Screening Hotline.
Question 10.3: True or false: This video **has been speeded up** for this experiment.
Question 10.4: True or false: In the video, the user works at a pharmacy.

III. Agent Descriptions

What follows is a description of the two agent types (i.e. human Robin and chatbot Robin) as they appear in the survey.

Agent 1: Chatbot Robin

What follows is a conversation between a user and Robin, a chatbot at Covid-19 Screening Hotline.

Notes:

- Chatbot is an example of Conversational AI; it is a piece of software that conducts conversations.
- This hotline is the first line of response for users who suspect that they might have contracted Covid-19 and guide them with the next recommended steps based on the screening results.

Agent 2: Human Robin

What follows is a conversation between a user and Robin, an agent at Covid-19 Screening Hotline.

Notes:

- Robin is a qualified staff member trained to be the first line of response for users who suspect that they might have contracted Covid-19 and guide them with the next recommended steps based on the screening results.

IV. Item Reliability

Item Reliability for Disposition to Trust

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
disp1	371	+	0.9121	0.8547	1.321433	0.8967
disp2	371	+	0.8736	0.7883	1.347731	0.9115
disp3	371	+	0.8954	0.8332	1.381292	0.9014
disp4	371	+	0.8747	0.7970	1.382499	0.9085
disp5	371	+	0.8359	0.7624	1.559297	0.9170
Test scale					1.39845	0.9244

Item Reliability for PIIT

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
piit1	371	+	0.9218	0.8604	1.687038	0.8673
piit2	371	+	0.8765	0.7633	1.687487	0.9034
piit3	371	+	0.8316	0.7035	1.889626	0.9201
piit4	371	+	0.9433	0.8983	1.641218	0.8549
Test scale					1.726342	0.9126

Item Reliability for Satisfaction

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
sat1	371	+	0.8938	0.8440	.4988004	0.9035
sat2	371	+	0.8740	0.8209	.5178276	0.9061
sat3	371	+	0.9150	0.8703	.4769782	0.9009
sat4	371	+	0.6506	0.5783	.6356902	0.9292
sat5	371	+	0.7179	0.6326	.589349	0.9240
sat6	371	+	0.8682	0.8135	.5215119	0.9069
sat7	371	+	0.8560	0.7904	.5112159	0.9095
Test scale					.5359105	0.9239

Item Reliability for Trust - Ability

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
ability1	371	+	0.8296	0.6733	1.053318	0.8957
ability2	371	+	0.9374	0.8294	.6071028	0.7581
ability3	371	+	0.9134	0.7995	.7390471	0.7809
Test scale					.7998227	0.8730

Item Reliability for Trust - Benevolence

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
benev1	371	+	0.9003	0.7703	1.489838	0.7917
benev2	371	+	0.9271	0.8258	1.317855	0.7383
benev3	371	+	0.8391	0.6513	1.862883	0.8979
Test scale					1.556859	0.8674

Item Reliability for Trust - Integrity

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
integ1	371	+	0.8882	0.7635	1.267844	0.9154
integ2	371	+	0.9318	0.8431	1.059161	0.8501
integ3	371	+	0.9405	0.8561	.9925184	0.8392
Test scale					1.106508	0.9097

Item Reliability for Trust in Agent

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
tworthy1	371	+	0.8982	0.8398	1.171151	0.9223
tworthy2	371	+	0.9307	0.8932	1.170244	0.9141
tworthy3	371	+	0.9399	0.9042	1.126268	0.9107
tworthy4	371	+	0.9325	0.8947	1.156819	0.9133
tworthy5	371	+	0.8153	0.6895	1.185452	0.9578
Test scale					1.161987	0.9379

Item Reliability for Persuasiveness

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
pursuasive1	371	+	0.8634	0.8032	.7117812	0.9117
pursuasive2	371	+	0.8880	0.8324	.6800619	0.9074
pursuasive3	371	+	0.8279	0.7573	.7369323	0.9176
pursuasive4	371	+	0.8937	0.8354	.6580433	0.9070
pursuasive5	371	+	0.8727	0.8109	.6901027	0.9103
pursuasive6	371	+	0.7923	0.6963	.7283267	0.9256
Test scale					.7008747	0.9268

Item Reliability for Likelihood of Use

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
use1	371	+	0.9626	0.9332	1.961876	0.9590
use2	371	+	0.9794	0.9621	1.870091	0.9507
use3	371	+	0.9175	0.8609	2.153049	0.9789
use4	371	+	0.9740	0.9517	1.87053	0.9538
Test scale					1.963886	0.9705

Item Reliability for Follow Advice

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
followup1	371	+	0.9071	0.8218	1.021066	0.8324
followup2	371	+	0.9038	0.8076	.992499	0.8394
followup3	371	+	0.9175	0.8492	1.050375	0.8240
followup4	371	+	0.7303	0.5715	1.431575	0.9197
Test scale					1.123879	0.8895

V. Demographic Distribution in the Four Conditions

	Gender		Total
	Male	Female	
Human and Mild	36	55	91
Human and Severe	47	45	92
Chatbot and Mild	49	44	93
Chatbot and Severe	51	44	95
Total	183	188	371

Chi Square=4.729, p=.193

	Age							Total
	18-24	25-34	35-44	45-54	55-64	66-74	75-84	
Human and Mild	1	22	28	18	16	6	0	91
Human and Severe	0	27	30	19	12	4	0	92
Chatbot and Mild	0	23	35	17	14	4	0	93
Chatbot and Severe	2	26	32	16	13	5	1	95
Total	3	98	125	70	55	19	1	371

Chi Square=9.556, p=.945

	Ethnicity				Total
	Asian	Black	Other	White	
Human and Mild	5	9	1	76	91
Human and Severe	7	4	5	76	92
Chatbot and Mild	6	8	4	75	93
Chatbot and Severe	11	2	1	81	95
Total	29	23	11	308	371

Chi Square=13.336, p=.148

Movie S1.

Video for Mild Symptoms: https://youtu.be/L4sjeuhULiw

Movie S2.

Video for Severe Symptoms: https://youtu.be/E-YOMDjJlVo

Data S1.

Data File: https://iu.box.com/v/JAMIADataExport

References

1. Thielsch MT, Meeßen SM, Hertel G. Trust and distrust in information systems at the workplace. PeerJ 2018;**6**:e5483
2. Trust Management — An Information Systems Perspective. Proceedings of the 27th European Conference on Information Systems (ECIS); 2019.
3. Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: A cross-discipline view of trust. Academy of management review 1998;**23**(3):393-404
4. McKnight DH. Trust in information technology. The Blackwell encyclopedia of management 2005;**7**:329-31
5. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology: An investigation of its components and measures. ACM Transactions on management information systems (TMIS) 2011;**2**(2):1-25
6. Thatcher JB, Perrewe PL. An empirical examination of individual traits as antecedents to computer anxiety and computer self-efficacy. MIS quarterly 2002:381-96
7. Melone NP. A theoretical assessment of the user-satisfaction construct in information systems research. Management science 1990;**36**(1):76-91
8. Doll WJ, Deng X, Raghunathan TS, Torkzadeh G, Xia W. The meaning and measurement of user satisfaction: A multigroup invariance analysis of the end-user computing satisfaction instrument. Journal of Management Information Systems 2004;**21**(1):227-62
9. Xiao L, Dasgupta S. Measurement of user satisfaction with web-based information systems: An empirical study. AMCIS 2002 Proceedings 2002:159
10. Baroudi JJ, Orlikowski WJ. A short-form measure of user information satisfaction: a psychometric evaluation and notes on use. Journal of Management Information Systems 1988;**4**(4):44-59
11. Gupta B, Dasgupta S, Gupta A. Adoption of ICT in a government organization in a developing country: An empirical study. The Journal of Strategic Information Systems 2008;**17**(2):140-54
12. Dennis AR, Robert Jr LP, Curtis AM, Kowalczyk ST, Hasty BK. Research note—trust is in the eye of the beholder: A vignette study of postevent behavioral controls' effects on individual trust in virtual teams. Information systems research 2012;**23**(2):546-58
13. Suka M, Yamauchi T, Yanagisawa H. Perceived effectiveness rating scales applied to insomnia help-seeking messages for middle-aged Japanese people: a validity and reliability study. Environmental health and preventive medicine 2017;**22**(1):69
14. De Vreede G-J. Collaborative business engineering with animated electronic meetings. Journal of Management Information Systems 1997;**14**(3):141-64
15. Nicolaou AI, McKnight DH. System design features and repeated use of electronic data exchanges. Journal of Management Information Systems 2011;**28**(2):269-304