

Brief Communications

User reactions to COVID-19 screening chatbots from reputable providers

Alan R. Dennis^{1*}, Antino Kim¹, Mohammad Rahimi², Sezgin Ayabakan²

¹Kelley School of Business, Indiana University, Bloomington, IN

²Fox 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 η^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 ^{***}
R²	0.269	0.216	0.193
Adjusted R²	0.234	0.178	0.154
F	5.363	5.101	8.434
Effect Sizes (Partial <i>eta</i>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***
R²	0.671	0.766	0.553	0.741	0.594
Adjusted R²	0.653	0.752	0.527	0.726	0.571
F	35.759	57.167	21.633	50.140	25.601
Effect Sizes (Partial <i>eta</i>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 Patients, 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