

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

---

12-2018

### Living with artificial intelligence – developing a theory on trust in health chatbots

Weiyu WANG

Keng SIAU

Singapore Management University, klsiau@smu.edu.sg

Follow this and additional works at: [https://ink.library.smu.edu.sg/sis\\_research](https://ink.library.smu.edu.sg/sis_research)



Part of the [Artificial Intelligence and Robotics Commons](#), [Databases and Information Systems Commons](#), and the [Health Information Technology Commons](#)

---

#### Citation

WANG, Weiyu and SIAU, Keng. Living with artificial intelligence – developing a theory on trust in health chatbots. (2018). *Proceedings of the 16th Annual Pre-ICIS Workshop on HCI Research in MIS, San Francisco, USA, 2018 December 13*. 1-5.

Available at: [https://ink.library.smu.edu.sg/sis\\_research/9402](https://ink.library.smu.edu.sg/sis_research/9402)

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [cherylids@smu.edu.sg](mailto:cherylids@smu.edu.sg).

# Living with Artificial Intelligence

## – Developing a Theory on Trust in Health Chatbots

**Weiyu Wang**

Missouri University of Science and Technology  
wwpmc@mst.edu

**Keng Siau**

Missouri University of Science and Technology  
siauk@mst.edu

### ABSTRACT

The current world is AI-filled and AI-fueled. Humans need to be able to live in harmony with AI. In this research, we aim to develop a theory on Trust between human and health chatbots. With the development of intelligent personal assistants, chatbots are becoming common and ubiquitous. Chatbots can behave as a conversational partner, complete information acquisition, and provide responses to inquiries. Chatbots have been widely used in the healthcare area, supporting physicians and assisting patients. The potential threats to privacy issues and the unpredictable performances of the chatbots hindered people's trust and adoption of the new technology. This research-in-progress paper discusses building a theory on trust in health chatbots. We will use a qualitative approach, Keeney's Value-Focused Thinking, in this theory-building exercise. Interviews with health chatbot vendors, users, and executives in medical centers that utilize health chatbots will be conducted.

### Keywords

Chatbots, Human-Intelligent Agent Interaction, Trust Building, Artificial Intelligence, Healthcare.

### INTRODUCTION

Since the term "Artificial Intelligence" (AI) was first coined in 1956 at a conference at Dartmouth College, AI has been developed and studied for about six decades. Most of the major breakthroughs appear over the last few years because of technology advancements and the availability of big data for machine learning. AI applications include self-driving cars, chatbots, and humanoid robots that apply face recognition, speech recognition, natural language processing, and other machine learning techniques. AI applications could be found in a myriad of industries, including finance, transportation, education, and healthcare. In the healthcare industry, AI applications range from monitoring patients' health conditions to diagnosis, disease rehabilitation, and assisting surgery. AI is propelling and fueling many new medical technology applications that will drastically change the way people interact with physicians and hospitals.

With the evolution of intelligent personal assistants, such as Apple's Siri, Amazon Alexa, Microsoft's Cortana, and

Google Assistant, which are supported by machine learning, chatbots do not exist in science fiction anymore. Chatbots can behave as a conversational partner, provide customer service, complete information acquisition, and respond to medical questions. Healthcare is critical to everyone. Although there are many medical questions that do not require the full attention of a physician, these questions will cause apprehension and trepidation for the concerned people if the questions are unanswered. Many people are searching online these days for medical advice and answers. This can be risky because of the uncertainty in the quality of information and the Internet is full of fake news and false information. Health Chatbot, which is supported and taught by physicians, can fill this gap to provide reliable and personalized answers. The developers of this software agents are working on making the conversations with chatbots more 'human-like' and enabling chatbots to be sensitive to the emotional background. The relationship between humans and intelligent agents are evolving. A study found that 63% of people would consider messaging an online chatbot to get in touch with a business or brand. However, as human-intelligent agents interaction developed, humans must trust that the intelligent agents will protect their interest and welfare. Inappropriate levels of trust may have negative consequences, such as misuse of the agents. To enhance the trust, chatbots can provide some positive reassuring cues, such as answering in a comforting human voice, that make users accept them socially.

Millions of chatbots have been sold to hospitals (Siwicki 2018). Health chatbots provide a variety of services such as empowering users to live independently and aiding seniors in breaking the loneliness. To benefit from the existing health chatbots, users need to register and provide some personal information, including some sensitive healthcare information. This, among other issues, has resulted in apprehension among the users.

Research has shown that trust plays an important role in directing human behaviors, including the adoption of an individual or an object. Trust is one of the fundamental pre-requisites for the prosperity of human societies (Fukuyama 1995). Research has demonstrated the significance of trust in different kinds of relationships, such as interpersonal relationships, human-social interactions (McKnight et al. 1998), e-commerce (Gefen 2000; Siau and Shen 2003), relationship in a virtual team

(Coppola et al. 2004), and human-technology interactions (Li et al. 2008). But few studies have focused on trust between human and artificial intelligent agents, such as chatbots. Intelligent agents are different from machines because artificial intelligent agents have the ability to learn and have the potential to perform like humans. This enables chatbots to convey more convincing humanness to users. Users who trust a chatbot may engage in a more personal conversation with and reveal more information to it. This research investigates the factors that affect trust between humans and intelligent agents and describes what are the differences with trust in other relations. The product of this theory building exercise will be a model of trust in health chatbots.

## LITERATURE REVIEW

### Definitions of Trust

Trust is “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party” (Mayer et al. 1995, p. 712). Trust means one believes in and is willing to depend on another party (McKnight et al. 1998). When considering online shopping environment, trust is defined as the belief that another party is benevolent, competent, honest, or predictable in a given situation and the willingness to depend on the party (Gefen et al. 2003). In the context of robots, trust is defined as the willingness of people to accept robot-produced information and follow the robot’s suggestions (Hancock et al. 2011). Several terms are used synonymously with trust, obfuscating the nature of trust. Cooperation, confidence, and predictability are among these terms. Trust can be asserted to mean the probability that a trustee’s performance will be beneficial to the trustor at a high enough level that the trustor will engage in some form of cooperation with the trustee (Gambetta 2000). The difference between trust and cooperation is that cooperation does not necessarily put a party at risk (Mayer et al. 1995). Trust is asserted as “the extent to which one is willing to ascribe good intentions to and have confidence in the words and actions of other people.” (Cook and Wall 1980, p. 39). To build trust, risk must be recognized and assumed in advance, but to have confidence, those risks are not necessary to be considered (Luhmann 2000). Many pieces of literature equate predictability with trust, however, predictability misses the willingness to take a risk in a relationship (Mayer et al. 1995). In another word, the risky situation is a prerequisite for trust arising.

There is a lack of clear differentiation between factors that are antecedents of trust, the outcome of trust, and trust itself. In general, trust is viewed as two elements: trusting beliefs (a set of specific beliefs dealing with benevolence, competence, integrity, and predictability) and trusting intention (the willingness of one party to depend on

another in a risky situation) (Siau and Wang 2018). In previous research, benevolence, integrity, and ability are categorized as perceived trustworthiness and are considered as the foundation of trust building (Mayer et al. 1995). Further, trust consists of two players: trustor and trustee (Siau and Shen 2003). When a trustor trusts a trustee, the trustor is vulnerable to the trustee’s actions.

Why would a party like to trust another party in a risky situation? It is affected by two main factors -- the attribute of the trustor and the attribute of the trustee. The attribute of the trustee refers to trustworthy (Mayer et al. 1995), including performance and predictability (Paravastu et al. 2014) in a human-technology relationship. In an interpersonal relationship, trustworthy indicates ability, integrity, and benevolence (Jian et al. 2000; Mayer et al. 1995). The attributes of trustor are studied in five main streams (Gefen et al. 2003; Li et al. 2008), including personality-based trust, cognition-based trust, calculative-based trust, institution-based trust, and knowledge-based trust. Personality-based trust, also mentioned as a disposition to trust, refers to the consistent tendency to be willing to depend on others (McKnight et al. 2000), including faith in humanity and trusting stance (Li et al. 2008). Recent research works use personality per se as factors that contribute to trust building (Bansal et al. 2016). Cognition-based trust refers to the cognitive cues, including the categorization process and control process (McKnight et al. 1998). One example of a categorization process is the categorization of trustees’ reputations (Siau and Wang 2018). Calculative-based trust refers to the calculation of pros and cons, and cost and benefit (Gefen et al. 2003). Institution-based trust means the sense of security one feels about a situation, including the regulations and guarantees provided by the institution (which the trustee belongs to) (Li et al. 2008). Knowledge-based trust refers to the first-hand knowledge and the prior experiences a trustor has toward the trustee (Kim 2012).

### Chatbots

The chatbot is a relatively new term. Abdul-Kader and Woods (2015) define chatbot as “a computer program that has the ability to hold a conversation with human using Natural Language Speech” (p.72), and “it is a computer program that mimics intelligent conversation” (p.73). Oh et al. (2017) use the term “chat bot” (also called chatterbot) rather than chatbot in their paper. They define chat bot as a computer program that “respond conversation to user via auditory or textual methods” (p373). They believe that chatbots should understand natural language dialogues and generate natural responses in a conversation. The common ground between the two definitions is the conversation using natural language in a dialogue. Since natural language dialogues are more natural than graphic-based interfaces, spoken dialogue systems are the primary interaction method with a machine (Nass and Brave 2015). Speech and conversation are the two most powerful forms of communication among humans. Researchers are always

ambitious to improve speech interaction between human and computer in order to simulate human-human speech interaction (Abdul-Kader and Woods 2015). Chatbots can assist in human-computer interaction and influence the behavior of users by asking questions and responding to the users' questions. In addition to speech ability, emotional intelligence is necessary for chatbots to function as a digital companion. A good chatbot must be able to identify and consider the emotional aspects of users and function in a way that is similar to the support provided by healthcare professionals, friends, and family. In other words, it should operate beyond rational aspects such as information provision and reasoning (Oh et al. 2017).

### Conceptual Foundation of Trust Building

Trust is dynamic. Before getting firsthand knowledge of the trustee, a trustor always decides whether to trust the trustee based on the trustor's disposition, the subjective norms, and the trustee's reputation (McKnight et al. 1998). This kind of trust is called initial trust (Siau et al. 2004). After getting firsthand knowledge, the trust will be built up in a gradual manner, requiring an ongoing two-way interaction (Kim 2012). This kind of trust is called ongoing trust or continuous trust (Siau and Shen 2003). Once the initial trust is built up, trustor will have some interactions with the trustee, such as conducting an online session or a dialog with health chatbots. If the experience is positive, the trust will be reinforced and enhanced, and continuous trust is enhanced. In addition, the five streams of trust antecedents mentioned above have an impact on the continuous trust. Among those antecedents, personality base, institution base, and cognition base are believed to have an impact on initial trust-building (Gefen et al. 2003; Kim 2012). Figure 1 depicts the basic trust-building process.

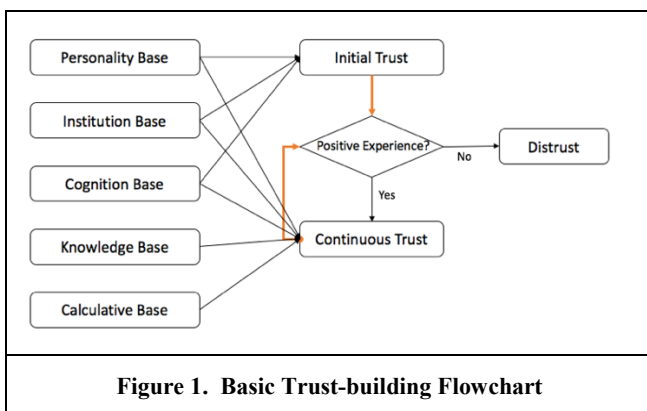


Figure 1. Basic Trust-building Flowchart

### Factors that Affect Trust Building in Chatbots

Several factors have been shown to affect trust building.

(i) Privacy concern is an increasingly important issue in an online environment (Bansal and Gefen 2010). Health chatbots, which rely on the online environment and

machine learning technology, may create privacy challenges.

(ii) Perceived Performance is “a set of beliefs about the capability of the software artifact to accomplish its designated purpose” (Paravastu et al. 2014, p.34). As an important component of trustworthiness, the ability is a key determinant of organizational trust (Mayer et al. 1995; McKnight et al. 1998).

(iii) Predictability is “the belief that the software artifact will do what it is claimed to do without adding anything malicious on top of it” (Paravastu et al. 2014, p.34). Health chatbots that are predictable are more dependable. Those chatbots can provide the users with expected outcomes.

(iv) Transparency and interpretability in health care are critical. Health information and outcomes should be understood, transparent, and interpretable by users. The machine learning algorithms (e.g., deep learning) and the chatbots may be black boxes (e.g., the decisions or recommendations made may not be explainable and the process traceable), and this lack of transparency is an issue.

(v) Subjective norms mean a “person’s perception that most people who are important to him think he should or should not perform the behavior in question” (Fishbein and Ajzen 1975, p.302). The impact of subjective norms is especially significant when users do not have sufficient understanding of a technology (Paravastu et al. 2014), such as AI or chatbots. Users tend to rely on subjective norms to help ascertain the assessments of the performance of an information system (Venkatesh and Davis 2000).

(vi) Familiarity and Experience have been studied in different contexts (Gefen and Straub 2004; Song 2007). In studying trust on chatbots, we will need to consider the familiarity and experience of the subjects on AI and chatbots.

(vii) Propensity and Personality are important factors in trust building too. The trustor's propensity to trust affect his/her trust for a trustee prior to the availability of information about the trustee (Mayer et al. 1995). Personality has a significant impact on initial trust-building (McKnight 1998) when firsthand knowledge and experience have not been obtained.

(viii) Nature of healthcare tasks (e.g., importance and complexity) will affect people’s willingness to follow an intelligent agent’s instruction (Salem et al. 2015).

### RESEARCH METHODOLOGY

This research aims to develop a theory of trust between human and health chatbots. According to Gregor (2006), there are five theory types: theory for analysis, theory for explanation, theory for prediction, theory for explanation and prediction, and theory for design and action. In information systems, design theories are intended to give

guidance to developers (Markus et al. 2002). However, allocating a theory building exercise to one class is not straightforward. Some subjective judgmental calls may be needed to determine which theory type it belongs. This research aims to propose a theory for human-intelligence agents' interaction that enhances trust between human and health chatbots. The objective of the research matches that of the design and action type. The criteria of design science include the "utility to users, the novelty of the artifact, and the persuasiveness of claims that it is effective" (Gregor 2006 p.629). This research will be using a qualitative approach to construct a theory to illustrate the trust relationship between humans and health chatbots. The trust model can be used by developers in the development of a chatbot that can be trusted by users.

According to Walls et al. (1992), an information system design theory includes three elements: a set of user requirements, a set of system features, and a set of principles deemed effective for guiding the process of development. To do this, we would interview health chatbots vendors, physicians that worked with health chatbots, users that interacted with health chatbots, and executives in the healthcare industry.

Keeney's Value-Focused Thinking (VFT) approach (Keeney 1992, 1994) will be utilized as the interviewing technique to solicit the factors. VFT, which is fundamentally about deciding what is important and how to achieve it, defines essentially what decision makers care about. Values are principles used for the evaluation (Keeney, 1992). Values that are of concern are made explicit by the identification of objectives. An objective is a statement of something that one *desires* to achieve (Keeney, 1992). An objective is characterized by three features: a decision context, an object, and a direction of preference. VFT not only uncovers hidden objectives, but also provides a systematic way of identifying relationships among the objectives. VFT has been utilized in several IS/IT research (e.g., Siau et al. 2004, Sheng et al. 2005, Nah et al. 2005,) and has been shown to be very helpful in determining the fundamental values and objectives that are usually hidden.

VFT involves the following three steps:

- develop an initial list of objectives and express all objectives in a common form
- structuring objectives to identify fundamental and means objectives
- building the means-ends objective network

## EXPECTED CONTRIBUTIONS AND CONCLUSIONS

Trust is crucial for the continuing evolution and adoption of health chatbots and other advanced AI applications. AI is impacting every aspect of our life and it is redefining the future of work and humanity. Healthcare is one area that is going to be heavily impacted by AI and AI has the potential to advance healthcare and reduce the costs of

healthcare. Health chatbots are common these days and their potential are tremendous. However, to achieve the potential, the trust issues need to be understood, managed and addressed.

This research aims to develop a theory on trust in health chatbots. The research will use the Keeney's Value-focused Thinking Approach. For academics, this research will identify the factors contributing to trust in health chatbots. As a research study in the stream of research to study the broader trust issues in AI, this research will provide the preliminary results and foundation for this stream of research. For practitioners, we expect the outcome of this study to provide insights into the design of health chatbots to foster trust between users and health chatbots.

## REFERENCES

1. Abdul-Kader, S. A., and Woods, J. C. 2015. "Survey on chatbot design techniques in speech conversation systems," *International Journal of Advanced Computer Science and Applications*, 6(7).
2. Bansal, G., and Gefen, D. 2010. "The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online," *Decision support systems*, 49(2), 138-150.
3. Bansal, G., Zahedi, F. M., and Gefen, D. 2016. "Do context and personality matter? Trust and privacy concerns in disclosing private information online," *Information & Management*, 53(1), 1-21.
4. Cook, J., and Wall, T. 1980. "New work attitude measures of trust, organizational commitment and personal need non-fulfilment," *Journal of Occupational and Organizational Psychology*, 53(1), 39-52.
5. Coppola, N. W., Hiltz, S. R., and Rotter, N. G. 2004. "Building trust in virtual teams," *IEEE transactions on professional communication*, 47(2), 95-104.
6. Fishbein, M., and Ajzen, I. 1975. "Belief, attitude, intention and behavior: An introduction to theory and research".
7. Fukuyama, F. 1995. "Trust: The social virtues and the creation of prosperity," *No. D10 301 c. 1/c. 2*, Free Press Paperbacks.
8. Gambetta, D. 2000. "Can we trust trust," *Trust: Making and breaking cooperative relations*, 13, 213-237.
9. Gefen, D. 2000. "E-commerce: the role of familiarity and trust," *Omega*, 28(6), 725-737.
10. Gefen, D., Karahanna, E., and Straub, D. W. 2003. "Trust and TAM in online shopping: An integrated model," *MIS quarterly*, 27(1), 51-90.

11. Gefen, D., and Straub, D. W. 2004. "Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services," *Omega*, 32(6), 407-424.
12. Gregor, S. (2006). The nature of theory in information systems. *MIS quarterly*, 611-642.
13. Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. 2011. "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, 53(5), 517-527.
14. Jian, J. Y., Bisantz, A. M., and Drury, C. G. 2000. "Foundations for an empirically determined scale of trust in automated systems," *International Journal of Cognitive Ergonomics*, 4(1), 53-71.
15. Keeney, R.L. 1992. "Value-focused Thinking", Harvard University Press, Massachusetts, Cambridge
16. Keeney, R.L. 1994. "Creativity in decision making with value – focused thinking", *Sloan Management Review*, Summer, pp. 33-41.
17. Kim, J. B. 2012. "An empirical study on consumer first purchase intention in online shopping: integrating initial trust and TAM," *Electronic Commerce Research*, 12(2), 125-150.
18. Li, X., Hess, T. J., and Valacich, J. S. 2008. "Why do we trust new technology? A study of initial trust formation with organizational information systems," *The Journal of Strategic Information Systems*, 17(1), 39-71.
19. Luhmann, N. 2000. "Familiarity, confidence, trust: Problems and alternatives," *Trust: Making and breaking cooperative relations*, 6, 94-107.
20. Markus, M. L., Majchrzak, A., & Gasser, L. 2002. "A design theory for systems that support emergent knowledge processes," *MIS quarterly*, 179-212.
21. Mayer, R. C., Davis, J. H., and Schoorman, F. D. 1995. "An integrative model of organizational trust," *Academy of management review*, 20(3), 709-734.
22. McKnight, D. H., Cummings, L. L., and Chervany, N. L. 1998. "Initial trust formation in new organizational relationships," *Academy of Management review*, 23(3), 473-490.
23. McKnight, D. H., Choudhury, V., and Kacmar, C. 2000. "Trust in e-commerce vendors: a two-stage model," *In Proceedings of the twenty first international conference on Information systems*, Association for Information Systems, pp. 532-536.
24. Nah, F., Siau, K., Sheng, H., "The Value of Mobile Applications: A Utility Company Study," *Communications of the ACM*, Vol. 48, No. 2, February 2005, pp. 85-90.
25. Nass, C.I. and Brave, S. 2015. "Wired for speech: How voice activates and advances the human-computer relationship", MIT Press Cambridge.
26. Oh, K. J., Lee, D., Ko, B., and Choi, H. J. 2017. "A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation," *In Mobile Data Management (MDM), 2017 18th IEEE International Conference on* (pp. 371-375). IEEE.
27. Paravastu, N., Gefen, D., and Creason, S. B. 2014. "Understanding trust in IT artifacts: An evaluation of the impact of trustworthiness and trust on satisfaction with antiviral software," *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 45(4), 30-50.
28. Salem, M., Lakatos, G., Amirabdollahian, F., and Dautenhahn, K. 2015. "Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust," *In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 141-148, ACM.
29. Sheng, H., Nah, F., Siau, K., "Strategic Implications of Mobile Technology: A Case Study Using Value-Focused Thinking," *Journal of Strategic Information Systems*, Vol. 14, No. 3, 2005, pp. 269-290.
30. Siau, K., and Shen, Z. 2003. "Building customer trust in mobile commerce," *Communications of the ACM*, 46(4), 91-94.
31. Siau, K., Sheng, H., Nah, F., and Davis, S. 2004. "A qualitative investigation on consumer trust in mobile commerce," *International Journal of Electronic Business*, 2(3), 283-300.
32. Siau K., and Wang, W. 2018. "Building Trust in Artificial Intelligence, Machine Learning, and Robotics," *Cutter Business Technology Journal*, Vol. 31, No. 2, pp. 47-53.
33. Siwicki, B. 2018. "Special Report: AI voice assistants have officially arrived in healthcare," Retrieved from <http://www.healthcareitnews.com/news/special-report-ai-voice-assistants-have-officially-arrived-healthcare>.
34. Song, J. 2007. "Trust in health infomediaries," *Decision support systems*, 43(2), 390-407.
35. Venkatesh, V., and Davis, F. D. 2000. "A theoretical extension of the technology acceptance model: Four longitudinal field studies," *Management science*, 46(2), 186-204.
36. Walls, J. G., Widmeyer, G. R., and El Sawy, O. A. 1992. "Building an Information System Design Theory for Vigilant EIS," *Information Systems Research* (3:1), pp. 36-5.