Towards a Model for Building Trust and Acceptance of Artificial Intelligence Aided Medical Assessment Systems

Lennart Seitz
Leuphana University Lüneburg
Sigrid Bekmeier-Feuerhahn
Leuphana University Lueneburg
Florian Bontrup
Leuphana University Lueneburg
Justus Wildt
Docyet
Krutika Gohil
Docyet

Cite as:
Seitz Lennart, Bekmeier-Feuerhahn Sigrid, Bontrup Florian, Wildt Justus, Gohil Krutika (2020), Towards a Model for Building Trust and Acceptance of Artificial Intelligence Aided Medical Assessment Systems. *Proceedings of the European Marketing Academy*, 49th, (64418)

Paper from the 49th Annual EMAC Conference, Budapest, May 26-29, 2020.

Towards a Model for Building Trust and Acceptance of Artificial Intelligence Aided Medical Assessment Systems

Abstract

This study aims to identify determinants for the emergence of trust in AI-based medical assessment systems consisting of chatbots and telemedicine. Existing studies have been failing to create a holistic understanding due to focusing on single trust antecedents. Our study closes this research gap by conducting semi-structured interviews and standardized questionnaires to identify relevant variables and their relationship to each other. Participants (n = 40) take part in a laboratory experiment interacting with a chatbot (vs. chatbot + human agent) for initial medical assessment. The first results indicate the importance of the chatbot's purpose and the transparency of underlying data base. Furthermore, communication patterns conveying uncertainty reduction are found to be more important than chatbot's social skills. The additional human expert complements the chatbot due to the possibility of more specific and detailed questioning and patients' wish of having a responsible person.

Keywords

Chatbots, healthcare, trust

Track

Services Marketing

1. Introduction

Possible applications of Artificial Intelligence (AI) based conversational agents, also known as chatbots, is not limited to customer services in e-commerce, but they are also deployed in more complex areas such as medical assessment or therapy planning (Denecke, Tschanz, Dorner, & May, 2019; Fitzpatrick, Darcy, & Vierhile, 2017). As digital health innovation continues to advance, it has initiated changes in providing care by changing the doctor-patient relationship with shared decision-making, communication, health management, and cost-effectiveness (Mesko, & Györffy, 2019). In this context, the acceptability of healthcare chatbots has been explored by a few studies (Morris, Kouddous, Kshirsagar, & Schueller, 2018; Ly, Ly, & Anderson, 2017) using specific measures of understanding, hesitancy, trust, and motivations. These studies revealed that 80% of the participants were very curious about using new technology to improve their health, and 65% perceived healthcare chatbots a good idea, indicating perceived belief and trust in chatbots as the agents (Nadarzynski, Miles, Cowie, & Ridge, 2019).

Given that people could envision themselves interacting with chatbots in the highly sensitive domain of health, chatbots used for initial assessment and treatment recommendations have to fulfill stricter requirements than those used for e-commerce. Even though technology will continue to proceed rapidly, the effect of AI in a complex system like healthcare can never be fully predicted, and it is not foreseeable that chatbots will work perfect soon (Kim, Coeira, & Magrabi, 2016). There are two approaches to addressing this: (1) enhancing the perceived quality of interaction with chatbots and (2) including an additional human expert agent (hybrid approach). This study aims to collect findings of user perception of such hybrid assistant systems that strike a beneficial balance between the effective use of automation and the integration of human experts. Particular attention will be paid to trust creation, as this is a key determinant in adopting new technology in highly sensitive environments.

2. Conceptual Background

2.1. Chatbots – digital artifact or social actor?

Chatbots are integrated on numerous websites offering the opportunity of fast customer support independently from service staff. Usually, users communicate via text messages with chatbots, which makes the conversation quite similar to interpersonal chat-mediated communication (e.g., WhatsApp). Although chatbots are an inanimate tool on websites

designed to allow users to inform themselves, they also show human-like communication patterns. Therefore, whether chatbot interaction feels similar to a real conversation or it is perceived as a source of self-information, should be discussed.

According to Social Response Theory, humans tend to respond socially to anything that displays human-like characteristics such as animals or technologies (Reeves, & Nass; 1996; Moon, 2000). Although chatbots are inanimate, they take the role of an interaction partner simulating human communication, which is why people may even attribute personality traits to them (Cassell, & Bickmore, 2000). Additionally, Social Presence Theory (Short, Williams, & Christie, 1976) postulates that communication media convey the social presence of another person. Considering chats are typically characterized by receiving and giving feedback immediately, such conversations can be categorized as being rich in terms of Media Richness Theory (Daft, & Lengel, 1984) and synchronous according to Media Synchronicity Theory (Dennis, Fuller, & Valacich, 2008). Although these theories have their origin in computer-mediated communication domains, the feeling of social presence is also relevant in interaction with AI (Biocca, Harms, & Burgoon, 2003). For example, studies have been shown that anthropomorphic characteristics, such as empathy expression, has positive effects on the perception of chatbots (Liu, & Sundar, 2018). Accordingly, the sociability of a chatbot is related to perceived trustworthiness (Heerink, Kröse, Evers, & Wielinga, 2010).

Nevertheless, there are also differences between human-chatbot-interaction and human-human-interaction. Although basic communication patterns appear to be quite similar, humans are less confident when communicating with chatbots (Hill, Ford, & Farreras, 2015). People react more frustratedly to virtual assistants compared to human agents, tend to make offensive remarks, and have lower confidence in artificial agents (Jenkins, Churchill, Cox, & Smith, 2007). A major problem is the insufficient ability of chatbots to maintain long-lasting and natural conversations (Lotze, 2016). Even state of the art chatbots are identified as such relatively quickly (Skjuve, Haugstveit, Folstad, & Brandtzaeg, 2019), and humans are aware of their irrational social behavior towards chatbots (Reeves, & Nass, 1996). Therefore, the extent to which interpersonal models fit describing the perception of chatbots is disupted (Gefen, Benbasat, & Pavlou, 2008; Riedl, Mohr, Kenning, Davis, & Heekeren, 2011).

The present study, therefore, locates the chatbot on a continuum between the human agent and self-information processes.

RQ1: How do people perceive the information received from a chatbot compared to information researched by themselves (e.g., on websites) vs. information received from an interaction partner (e.g., a doctor) and where are the differences?

2.2. Generating trust in health care assessment systems

The well-established trust model by Mayer, Davis, and Schoorman (1995) defines trust as "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" (p. 712). Due to their lack of medical knowledge, patients usually do not have the opportunity to control the quality of the doctor's work and must, therefore, rely on his assessment. Concerning the principal-agent problem, the doctor-patient relationship is an asymmetrical constellation in which patients are situated in an inferior position and must have confidence in the doctor (Scott, & Vick, 1999). Since illnesses always represent a threat to health and perceived risk moderates the willingness to trust, trust plays an important role in medical assessment. According to Hillen, Postma, Verdam, and Smets (2017), a patient's trust is "the optimistic acceptance of a vulnerable situation in which the patients believe the physician cares for his or her interests" (p. 855). Therefore, relevant dimensions of creating trust between doctor and patient are honesty, competence, caring, fidelity, and confidentiality (Hall et al., 2002; Hillen et al., 2017).

Trust does not only arise exclusively in interpersonal relationships, but people can also ascribe the trustee role to a computer (Riegelsberger, Sasse, & McCarthy, 2005). According to literature, the dimensions and predictors of the emergence of trust in technology ("e-trust"), however, differ from those described in interpersonal models. Söllner, Hoffmann, Hoffmann, Wacker, and Leimeister (2012) define performance, process, and purpose as important predictors for trust in a technological artifact. Of particular importance is the user's need to understand how the technology works and how it came to its output. Therefore, transparency and justification are significant e-trust antecedents (Taddeo, 2009; Zhao, Benbasat, & Cavusoglu, 2019).

Considering the use of chatbots for medical assessment is a rather novel technology, and people have little experience in interacting with virtual assistants (Hill et al., 2015), uncertainty and risk are highly salient (McKnight, Choudhury, & Kacmar, 2002). Perceived trust can, therefore, enhance the probability of technological innovation being accepted and adopted. This study investigates whether trust in chatbots used for the preliminary medical

assessment can be better explained by (1) interpersonal trust models or (2) e-trust models and how the requirements for chatbots differ from those for doctors.
RQ2: How can chatbots deployed for the preliminary medical assessment create trust?

2.3. Relevance of usability for generating trust in communication with chatbots

The perceived quality of experience with software is mainly influenced by usability, which describes how pleasant a person experiences the use of a technological artifact. While the Technology Acceptance Model (TAM) defines usability briefly as a function of perceived usefulness and perceived ease of use (Davis, 1989), other models consider further dimensions such as the ease of learning, general satisfaction and screen layout (e.g., Chin, Diehl, & Norman, 1988; Lund, 2001). While many established usability measurement scales focus on handling aspects and design, there are specific models for evaluating the usability of chatbots. Quarteroni and Manandhar (2007) emphasized the relevance of dialogue design and mainly integrated items for capturing the perception of information exchange between a chatbot and humans into their scale. Claessen, Schmidt, and Heck (2017) also defined usefulness and ease of use as successful information exchange and added the dimensions of trust and fun to their model.

Wang & Benbasat (2005) illustrated that there are numerous relationships between the dimensions of trust, perceived ease of use, perceived usefulness, and intention to adopt for the domain of recommendation agents. However, at the time of the experiment, such agents were not able to communicate as progressively and coherently as they do today. In their research plan, Gefen et al. (2008) already mentioned the potential of signaling through text content to generate e-trust without referring explicitly to the chatbot context. Thus, it remains to be answered to what extent communication patterns like expressing empathy or use of repetition affect perceived usability and trust in highly sensitive environments. Since the involvement of a human agent also has positive effects on the acceptance of virtual agents for healthcare domains (Söllner, & Königstorfer, 2019), the potential of enhancing perceived usability and trust by including a human expert agent will also be investigated.
RQ3: How do people experience the information exchange with chatbots in highly sensitive environments such as healthcare, and how can the experienced usability be improved by (1) specific communication patterns and (2) the addition of a human agent?

3. Method

3.1. Study design and sample

Our sample will consist of healthy participants (n = 40) with no history of psychiatric or neurological disease. Each participant is randomly assigned to one of the two experiments (chatbot interaction only vs. chatbot + human agent) and one of two scenario conditions (scenarios: mild severity disease vs. moderate severity disease) while ensuring equal chance of receiving any of the scenarios or experiments. All participants are naive to the experimental design, and they give written informed consent before the experiment began.

3.2. Scenarios

The mild severity disease scenario describes the symptoms of a cold, and the medium severity disease scenario depicts a bladder infection. Mild and moderate severity diseases are further classified into five different triage levels, which gives treatment recommendations to the participants depending on the severity of the disease. A pre-manipulation check (n = 23) was conducted to compare both scenarios concerning the dimensions imaginability, anxiety creation, and perceived severity of the disease using multiple items and a seven-point Likert scale. Results show that both diseases were well imaginable (M_{mild} = 5.81, SD_{mild} = 1.50; $M_{moderate}$ = 6.15, $SD_{moderate}$ = 1.06) and the bladder infection was perceived as being more worrying (M_{mild} = 1.86, SD_{mild} = 1.14; $M_{moderate}$ = 3.14, $SD_{moderate}$ = 1.67, $t(16) = -1.766$, $p < .05$) and more severe (M_{mild} = 3.33, SD_{mild} = 0.98; $M_{moderate}$ = 4.48, $SD_{moderate}$ = 1.38, $t(16) = -1.915$, $p < .05$) when compared to the cold.

After the initial assessment by the chatbot, one-half of the participants from both disease conditions are connected via mobile phone to a medical doctor (agent) using the telemedicine feature of the software. The medical doctor has access to the conversation between the chatbot and human using a specialized code given by the chatbot. The doctors are supposed to answer open questions and to give further treatment recommendations to participant.

3.3. Procedure and data collection

Participants are invited to take part in a laboratory experiment in which they interact with a chatbot used for initial medical assessment. In order to get unaffected impressions of the attitudes towards chatbots, an open pre-interview is conducted before starting the experiment. This pre-interview also records the self-information behavior on health issues (e.g., websites) and captures how self-information is experienced. Afterwards, the scenario description is handed out to participants, and participants are asked to interact with the chatbot to receive an

initial medical assessment. In its assessment, the chatbot shows approximately two to four possible diseases ranked by their probability, informing the patient about typical symptoms, a short explanation, and treatment recommendations. Corresponding to the study design, one-half of the participants are connected to telemedicine after the initial assessment. After the experiment, a semi-structured interview takes place and a standardized questionnaire is handed out capturing important dimensions of perceived trust, usability and communication. During the experiment, we also use an eye-tracker and facial scanner to collect further data about experiences and perceived usability.

4. Preliminary Results

The preliminary analysis procedure was based on Qualitative Content Analysis given by Mayring (2010). In first step, the interviews were transcribed, and relevant statements were identified. Second, similar paraphrases were summarized in key statements that form the basis for the formation of inductive categories. In addition, categories were formed deductively from the conceptual background, which was compared with and supplemented by inductive categories. The categories will be adjusted continuously as required, and the data material is reviewed repeatedly to develop an adequate final coding guideline. Due to a large number of interviews and in accordance with Mayring (2010), the final coding guideline will be set after 20% of the material (eight interviews). Simultaneously, the adequacy of the interview guidelines will be evaluated and optimized if necessary.

After reviewing four interviews, the first recurring statements and important concepts were identified. The positioning of chatbots between self-information sources and doctors was relatively balanced with a tendency towards the doctor. The reasons for a closer positioning towards the doctor were the following: (1) active participation, (2) systematic data analysis, and (3) trust in the accuracy of algorithms. The advantages of chatbots over web-based sources were (1) better data structure, and (2) a more trustworthy database. In contrast, a human doctor was perceived more trustworthy since he/she (1) had a comprehensible qualification, and (2) was able to respond to specific queries individually.

Asked for relevant aspects for the emergence of trust, participants mentioned antecedents of e-trust models more frequently than interpersonal trust dimensions. In particular, participants emphasized the importance of (1) the institution in the background, and (2) the sources of data. These reductions were associated with the IT-trust dimensions (1) purpose, and (2) transparency. Accordingly, participants also highlighted the importance of the institution’s intentions. Thus, the institution’s purpose might have been related to the

interpersonal trust dimensions of perceived benevolence and integrity as the chatbot can be perceived as a representative of the organization. Since the algorithm's accuracy may determine perceived competence, transparency of data sources may be linked to perceived ability. To further increase transparency, participants wanted to understand the path of data analysis and decision-making. Thus, justification may also be a major antecedent for creating trust in chatbots.

Besides transparency, communication patterns and usability appeared also to be related to the trustworthiness of chatbots. Participants frequently expressed concerns about how well the chatbot understood their entries. Several statements suggested that this uncertainty could be reduced by (1) using preset response options instead of free text entries, or (2) a short repetition by the chatbot after each message. However, preset response options would reduce (1) the perceived naturalness, and (2) the perceived specificity of the conversation, which may reduce trust and perceived flexibility. Furthermore, participants had low demands on the chatbot's social skills. Underlying reasons could be: (1) awareness that chatbots, obviously, are not alive, and (2) the superior desire for a fast and accurate assessment. In highly sensitive and risky environments, clear and reliable messages could be advantageous. Although participants expect the chatbot to adhere to basic communication and politeness rules, empathic reactions are denoted as inauthentic. The dialogue design of chatbots used for the medical assessment can, therefore, differ from those used for hedonic purposes. Perceived authenticity may also affect several outcome variables, such as perceived competence or trust.

The attitudes towards the integration of an additional human agent were very positive. However, it is important that participants (1) must know with whom they are connected, and (2) that the telemedical is a professional. The advantages of a supporting human agent were: (1) the possibility of specific questioning, (2) obtaining a second opinion, and (3) abdication of responsibility. For patients, it seemed to be important to have a person responsible in case of an incorrect assessment. Interestingly, trust in the human agent is only higher if he or she explicitly gives reasons for disagreeing with the chatbot's assessment or in cases of specific questions.

References

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Towards a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & Virtual Environments*, 12(5), 456–480.

Cassell, J. & Bickmore, T. (2000). External manifestation of trustworthiness in the interface. *Communications of the ACM*, 43(12), 50–56.

Chin, J. P., Diehl, V. A., & Norman, L. K. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In Soloway, E., Frye, D., & Sheppard, S. B. (eds.), *Proceedings of the ACM CHI 88 Human Factors in Computing Systems Conference* (pp. 213–218). Washington, ACM CHI.

Claessen, V., Schmidt, A., & Heck, T. (2017). Virtual assistants – A study on the usability and user perception of customer service systems for e-commerce. In Gäde, M., Trkulja, V., & Petras, V. (eds.), *Proceedings of the 15th International Symposium of Information Science* (pp. 116–130). Berlin, ISI.

Daft, R. L., & Lengel, R. H. (1984). Information richness: A new approach to managerial behavior and organizational design. *Research in Organizational Behavior*, 6, 191–233.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.

Denecke, K., Tschanz, M., Dorner, T. L., & May, R. (2019). Intelligent conversational agents in healthcare: Hype or hope? *Studies in health technology and informatics*, 259, 77–84.

Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, tasks, and communication processes: A theory of media synchronicity. *MIS Quarterly*, 32(3), 575–600.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Mental Health*, 4(2), e19.

Gefen, D., Benbasat, I., Pavlou, P. (2008). A research agenda for trust in online environments. *Journal of Management Information Systems*, 24(4), 275–286.

Hall, M. A., Zheng, B., Dugan, E., Camacho, F. Kidd, K. E., Mishra, A., & Balkrishnan, R. (2002). *Medical Care Research & Review*, 59(3), 293–318.

Heerink, M., Kröse, B., Evers, V., Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: the Almere Model. *International Journal of Social Robotics*, 2(4), 361–375.

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, 245–250.

Hillen, M. A., Postma, R.-M., Verdam, M. G. E., & Smets, E. M. A. (2017). Development and validation of an abbreviated version of the trust in oncologist scale–the trust in oncologist scale–short form (TiOS-SF). *Support Care Cancer*, 25, 855–861.

Jenkins, M.-C., Churchill, R., Cox, S., & Smith, D. (2007). Analysis of user interaction with service oriented chatbot systems. In Jacko, J. (ed.). *Human Computer Interaction*. *HCI Intelligent Multimodal Interaction Environments* (pp. 76–83). Berlin, Heidelberg: Springer Verlag.

Kim, M. O., Coeira, E., & Magrabi, F. (2016). Problems with health information technology and their effects on care delivery and patient outcomes: A systematic review. *Journal of the American Medical Informatics Association*, 24(2), 246–250.

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.

Lotze, N. (2016). *Chatbots. Eine linguistische Analyse [Chatbots. A linguistic analysis]*. Frankfurt a.M.: Peter Lang. (in German).

Lund, A. (2001). Measuring usability with the USE questionnaire. *Usability Interface*, 8(2), 3–6.

Ly, K. H., Ly, A.-M., Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions*, 10, 39–46.

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.

Mayring, P. (2010). Qualitative Inhaltsanalyse [Qualitative content analysis]. In Mey, G., & Mruck, K. (eds.) *Handbuch Qualitative Sozialforschung in der Psychologie* (p. 601–613). Wiesbaden, VS Verlag. (in German).

McKnight, D. H., Choudhury, V., Kacmar, C. (2002). Developing and validating trust measures for e-commerce: an integrative typology. *Information Systems Research*, 13(3), 334–359.

Mesko, B., Györffy, Z. (2019). The rise of the empowered physician in the digital health era: viewpoint. *Journal of Medical Internet Research*, 21(2): e12490.

Morris, R. R., Kouddous, K., Kshirsagar, R., Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *Journal of Medical Internet Research*, 20(6): e10148.

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research*, 26(4), 323–339.

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital Health*, 5, 1–12.

Quarteroni, S., & Manandhar, S. (2007). A chatbot-based interactive question answering system. In Artstein, R., & Vieu, L., *Proceedings on the 11th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 83–90). Rovereto, Decalog.

Reeves, B., Nass, C. (1996). *How people treat computers, television, and new media like real people and places.* Standford: CSLI Publications.

Riedl, R., Mohr, P., Kenning, P., Davis, F. D., & Heekeren, H. (2011). Trusting humans and avatars: Behavioral and neural evidence. In *Proceedings of the 32nd Conference on Information Systems* (7). Shanghai, ICIS.

Riegelsberger, J., Sasse, M. A., McCarthy, J. D. (2005). The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies*, 62, 381–422.

Scott, A., & Vick, S. (1999). Patients, doctors and contracts: an application of principal-agent theory to the doctor-patient relationship. *Scottish Journal of Political Economy*, 46(2), 111–134.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* London: John Wiley & Sons.

Skjuve, M., Haugstveit, I. M., Folstad, A., & Brandtzaeg, P. B. (2019). Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human–chatbot interaction. *Human Technology*, 15(1), 30–54.

Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J. M. (2012). Understanding the formation of trust in IT artifacts. In *Proceedings of the 33rd Conference on Information Systems* (11). Orlando, ICIS.

Söllner, M., & Königstorfer, J. (2019). Individuals' adherence to recommendations for the treatment of diseases depending on medical doctors' use of artificial intelligence. In *Proceedings of the 48th Annual Conference of the European Marketing Academy*. Hamburg, EMAC.

Taddeo, M. (2009). Defining trust and e-trust: from old theories to new problems. *International Journal of Technology and Human Interaction*, 5(2), 23–35.

Wang, W., & Benbasat, I. (2005). Trust in and adoption of online recommendation agents, *Journal of the Association for Information Systems*, 6(3), 72–101.

Zhao, R., Benbasat, I., & Cavusoglu, H. (2019). Do users always want to know more? Investigating the relationship between system transparency and users' trust in advice-giving systems. In *Proceedings of the 27th European Conference on Information Systems*. Stockholm, & Uppsala, ECIS.