THE ADVANCE OF DIAGNOSIS CHATBOTS: SHOULD WE FIRST AVOID DISTRUST BEFORE WE FOCUS ON TRUST?

Lennart Seitz  
*Leuphana University, lennart.seitz@leuphana.de*

Julia Woronkow  
*Leuphana University, julia.woronkow@leuphana.de*

Sigrid Bekmeier-Feuerhahn  
*Leuphana University, bekmeier@leuphana.de*

Krutika Gohil  
*DOCYET, krutika.gohil@docyet.com*

Follow this and additional works at: [https://aisel.aisnet.org/ecis2021_rip](https://aisel.aisnet.org/ecis2021_rip)

**Recommended Citation**  
[https://aisel.aisnet.org/ecis2021_rip/9](https://aisel.aisnet.org/ecis2021_rip/9)

This material is brought to you by the ECIS 2021 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2021 Research-in-Progress Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
THE ADVANCE OF DIAGNOSIS CHATBOTS: SHOULD WE FIRST AVOID DISTRUST BEFORE WE FOCUS ON TRUST?

Research in Progress

Lennart Seitz, Leuphana University Lüneburg, Germany, lennart.seitz@leuphana.de
Julia Woronkow, Leuphana University Lüneburg, Germany, julia.woronkow@leuphana.de
Sigrid Bekmeier-Feuerhahn, Leuphana University Lüneburg, Germany, bekmeier@leuphana.de
Krutika Gohil, Docyet, Leipzig, Germany, krutika.gohil@docyet.com

Abstract

Technological innovations like AI-based diagnosis chatbots will pave the way for a fundamental change in providing healthcare. While previous research mostly highlights the importance of trust in adopting new technologies, there is a lack of research on distrust. We argue that the avoidance of distrust towards novel technologies is crucial since negative emotions during initial interactions may harm subsequent trust-building processes. Therefore, our research aims to contribute to better understand the differences of trust and distrust towards diagnosis chatbots. To gain first deep insights into this new research area, we conducted a pilot usability study applying an innovative qualitative approach. First findings indicate that humans experience affect-based distrust and cognition-based trust towards diagnosis chatbots, while the opposite is true for doctor-patient-trust. We further argue that distrust goes beyond the absence of trust as it is triggered actively by negative emotions that may inhibit the adoption of new technologies.

Keywords: Trust, Distrust, Chatbot, Healthcare

1 Introduction

Driven by the ongoing digitization of numerous sectors, autonomous conversational agents' popularity has increased significantly in recent years (Araujo, 2018). Conversational agents are virtual assistants who can communicate with users via techniques like natural language processing to provide several kinds of services. One form of such agents are so-called chatbots using text messages to interact with humans (Feine et al., 2019). Among the various application areas is the healthcare domain where chatbots are used for diagnosis, elderly care, therapy, self-management, and prevention, to name just a few (Hauser-Ulrich et al., 2020; Montenegro et al., 2019). Their use facilitates cost and time-saving medical support and the provision of guided and reliable advice when integrated into existing health systems appropriately (Denecke et al., 2019; Powell, 2019). Furthermore, various studies refer to the positive health-related outcomes of using chatbots like desirable behavior changes or reduced depression symptoms (Bickmore et al., 2013; Fitzpatrick et al., 2017). Despite the numerous advantages, the implementation of health chatbots faces specific challenges due to the technology's novelty and the context's sensitivity. Especially in the severe situation of a medical consultation, patients must show a certain level of trust towards chatbots as it is a significant predictor for the willingness to adopt new technologies (Gefen et al., 2003). While much interdisciplinary research...
focuses on the ubiquitous concept of trust, distrust is considered less (Cho, 2006). Although both concepts are related, several authors argue that trust and distrust should be considered distinctively (Lewicki et al., 1998; McKnight et al., 2004). Since there are only a few studies addressing trust towards diagnosis chatbots (e.g., Laumer et al., 2019) and distrust in healthcare is barely conceptualized (Dean et al., 2017), our research aims to understand better the differences between the drivers of trust and distrust towards diagnosis chatbots from a patient's perspective. We argue that the avoidance of negative emotions induced by distrust is crucial for a successful implementation of health chatbots since negative emotions have a stronger impact on one's mental state than positive ones (McKnight and Chervany, 2001a; Vaish et al., 2008). To get first profound insights into trust- and distrust-building processes, we have taken advantage of a qualitative research design. We applied an innovative approach for collecting and analyzing data by mixing diverse methods allowing the participants to reflect deeply on their thoughts and experiences during the interaction with a diagnosis chatbot. Preliminary results and future steps are presented.

2 Related Research

2.1 Trust and distrust - two sides of the same coin?

In common parlance, the terms trust and distrust are often considered opposites as even dictionaries define distrust as the absence of trust (McKnight and Chervany, 2001b). However, there has been a broad scientific debate about the relation between both concepts over the past decades. While some scholars consider trust and distrust as opposites of a single continuum (Rotter, 1980), the majority of research assumes that they are indeed distinct concepts associated with different cognitive states and behavioral outcomes (McKnight and Chervany, 2001b). In the following, we will provide an overview of differences and similarities of the conceptualization of trust and distrust in literature.

The interdisciplinary well-researched concept of trust is seen as a fundamental concept in interpersonal relationships as it serves to reduce complexity in situations of uncertainty (Luhmann, 1979). Due to its omnipresence in several research domains, it is barely attainable to find a common definition for this multidimensional concept (McKnight and Chervany, 2001b). Regardless of whether it is considered a personality trait, a social structure or a calculative assessment, theorists generally agree that trust entails a psychological state of perceived risk and vulnerability (Kramer, 1999; McAllister, 1995). In this sense, trust can be defined as the trustor's optimistic expectation that another party will act in one's best interest accompanied by positive mental states like hope and confidence (Cho, 2006; Mayer et al., 1995). In how far trust emerges, depends on the trustor's propensity to trust, the trustor's beliefs concerning the trustee's trustworthiness as well as on situational and environmental circumstances (Kramer, 1999; McKnight and Chervany, 2001b). Trust can either be based on cognition when it emerges for good reasons or it can be based on affect when it is based on feelings and emotions (Lewis and Weigert, 1985). Cognition-based trust is typically associated with beliefs about a trustee's ability to perform a specific task and often occurs in professional exchange relationships. In contrast, affect-based trust entails general positive attitudes towards a person and is related to value- and morality-related dimensions like benevolence and warmth. Thus, affect-based trust typically occurs in relationships with strong emotional bonds (Chua et al., 2008; McAllister, 1995; Morrow et al., 2004). Trust, either triggered cognitively or affectively, evokes trusting intentions meaning that the trustor is willing to rely on the trustee's words or actions. Subsequently, these intentions manifest in actual trust-related behavior like securely depending on the trustee without the desire to control his or her actions (Mayer et al., 1995; McKnight and Chervany, 2001b).

In the literature on distrust, there is ambiguity regarding its relatedness to trust. Even if both concepts are considered distinctive, authors often use similar terms for their definitions. For instance, Lewicki et al. (1998) define distrust as the "confident negative expectation of another's conduct" while they define trust as the "confident positive expectation of another's conduct" (p. 439). Also, researchers often use equal constructs for the conceptualization of distrust, e.g., distrusting beliefs are formed through the evaluation of a trustee's competence and benevolence (Lee et al., 2015; McKnight and
Chervany, 2001b). Despite fundamental similarities, authors address qualitative differences to separate distrust from trust. First, they argue that distrust goes beyond the mere absence of trust as it is the active negative belief that a trustee will act with harmful intentions (Grover, 1994). Second, these negative beliefs are accompanied by emotions like suspicion, doubt or fear. Thus, the psychological state and neural correlates of distrust are associated with distinctive emotions (Cho, 2006; Dimoka, 2010). Under this assumption, a lack of trust does not necessarily result in distrust, just as the absence of distrust is not sufficient to evoke trust. Accordingly, the positive predictors of trust need not to be negative predictors of distrust and vice versa (Lee et al., 2015; Lewicki et al., 1998). Also, the consequences of distrust differ from those of low trust as negative emotions are more destructive than missing positive ones (Kramer, 1999; McKnight et al., 2004). Consequently, distrust triggers deeper cognitive processing and even avoidance behavior (Harding, 1993; Kramer, 1999).

Although distrust may not be the opposite of trust, it can be seen as a functional equivalent as it is the least desired and more complex way to deal with uncertainty (Luhmann, 1979). Therefore, we will consider trust and distrust as distinct but related psychological concepts that manifest in different cognitive and affective states entailing different antecedents and consequences. Accordingly, we assume that both concepts can indeed co-exist (Komiak and Benbasat, 2008; Lewicki et al., 1998).

### 2.2 Trusting and distrusting diagnosis chatbots

Even though trust and distrust are concepts to describe interpersonal phenomena, especially trust has frequently been adapted to IS research to investigate its role in interactions with several kinds of software systems (Söllner et al., 2012; Wang and Benbasat, 2008). Thereby, it is generally assumed that humans tend to apply interpersonal heuristics to their interactions with inanimate entities (Reeves and Nass, 1996). This may especially hold true for interactions with AI-based information systems as they imitate human intelligence and show immediate responsiveness. Furthermore, their capabilities of independent decision-making and their potential to outperform humans distinguishes AI fundamentally from prior generations of information systems (Glikson and Woolley, 2020; Jain et al., 2020). Due to AI's huge societal impact and complexity, users are faced with high levels of uncertainty making a certain level of trust indispensable (Glikson and Woolley, 2020; Lee and See, 2004).

The general adoption of interpersonal trust dimensions to IS research is discussed controversial, although there is empirical evidence for their applicability (Gefen et al., 2008). First, technological tools have no free will or malintentions raising the question in how far users form trusting beliefs like benevolence or integrity (Friedman et al., 2000; Söllner et al., 2012). Researchers thus suggest to conceptualize trusting beliefs towards technology by the equivalent dimensions performance, purpose and process (Lee and See, 2004; Söllner et al., 2012). Specifically, the provision of transparency is crucial for trust-building towards complex AI systems as it enables users to understand its inner workings (Al-Natour et al., 2010; Cramer et al., 2008). Second, trust between technological tools and human beings is not reciprocal but asymmetric. However, since AI-driven agents show several social characteristics, users may build up a certain kind of emotional attachment. Thus, affect-based trust may be triggered by anthropomorphic cues while cognition-based trust emerges through reliability beliefs and transparency (Glikson and Woolley, 2020; Wang and Siuau, 2018).

Regarding diagnosis chatbots in particular, the crucial role of trust becomes apparent considering the situation's sensitivity and the unique role of trust in doctor-patient-relationships (Anderson and Dedrick, 1990; Hall et al., 2001). One hurdle for trust-building towards diagnosis chatbots is the patients' need for interpersonal competencies like empathy and caring (Thom and Campbell, 1997). The high need for emotional intelligence distinguishes the provision of healthcare from other service contexts raising the question about users' expectations concerning emotional trust from diagnosis chatbots. As researchers argue, too much human-likeness could even lead to a sense of eeriness thus fear and distrust (Glikson and Woolley, 2020; Stein and Ohler, 2017). Furthermore, the acceptance of diagnosis chatbots may also be inhibited by the high societal renown of physicians and identity threats (LaRosa and Danks, 2018; Reis et al., 2020).
Since there are many approaches and conceptualizations of interpersonal and technology trust, we assume that fundamental determinants of trust development towards diagnosis chatbots may be quite similar. In contrast, to the best of our knowledge, there is almost no research that considers distrust-building towards diagnosis chatbots separately. Given the presented background and the scarce conceptualization of medical distrust (Dean et al., 2017), we argue that the understanding of distrust towards diagnosis chatbots should be a logical first step before focusing on the enhancement of trust.

3 Research Design

Considering that determinants of trust and distrust are often presented on more general terms in literature, our goal was to detect and distinguish the antecedents of trust and distrust more precisely. Since the emergence of trust and distrust is driven by individual experiences and expectations (Komiak and Benbasat, 2008), our research focuses on the users' subjective impressions during the interaction with a diagnosis chatbot. As quantitative research explains social behavior typical to a narrower extend, a qualitative approach fits our open research question best (Flick et al., 2012; Lamnek, 2010). Instead of adopting a specific theory, we decided to use a more open research process to get a broad first impression and understanding of this novel research field (Glaser and Holton, 2004). To get deep insights into perception processes and decision structures, we combined different qualitative approaches and various methods to holistically study the following research question:

What are the determinants of trust- and distrust-building processes towards diagnosis chatbots, and where are the differences between them?

3.1 Data collection procedure and sample

For our pilot study, we invited eight graduate, and undergraduate students enrolled at a German university through a call on social media to a laboratory experiment in which they interacted with a diagnosis chatbot. Instead of focusing on quantity, we followed usability research recommendations, which postulates that collecting deep insights can be more revealing than statistical analysis. We orientated on recommendations to invite between five and twelve participants (Macefield, 2009; Nielsen and Landauer, 1993). Each subject participated voluntarily in a guided part-narrative single-interview, which we supplemented with brainstorming, think-aloud, participatory observation, and a structure laying method. We documented the process in several ways: voice recording, transcription, pictures of visualizations, field notes, and memos. Since we adopted various methods, we will briefly discuss the research design to provide transparency and traceability (see Figure 1).
We started with (1) a warm-up to create an unconstrained atmosphere and continued with (2) a simple brainstorming asking about recent experiences with chatbots to develop an understanding of the participants' attitudes towards chatbots and technology. Critical statements were documented on a whiteboard as support for the following interview. Subsequently, we proceeded with (3) the actual chatbot interaction. We used a diagnosis chatbot from an mHealth provider that captures symptoms in a guided dialogue to give a preliminary medical assessment. Ahead of the interaction, we handed out a fictitious scenario to the participants, instructing to put themselves into the described situation of a sick patient who feels a strong back and neck pain, a slight headache, and a tingling in the right thumb. During the interaction with the chatbot, we asked the participants to think-aloud. We thereby aimed to develop a common ground and an understanding for the participant's inner discussion to observe trust- and distrust-building processes (Cooke, 2010; Van den Haak et al., 2003). Concurrently, the interviewer documented statements made by the participant on index cards, which were then pinned on a board. After the interaction, we (4) discussed what has been said and compared it to the prior statements revealed in the brainstorming to come to a deeper reflection. In the next step, we (5) asked more specific questions about the emergence of trust and distrust. Conclusively, the participant was asked to (6) assign all generated index cards on the board to the concept of trust or distrust (low and high). Afterwards, we discussed the participant's allocation procedure and reasons for the allocation.

### 3.2 Data validity

Due to the study's explorative character and the small sample size, conventional data validation procedures are hardly applicable. As an alternative strategy, we combined multiple methods to achieve triangulation (Flick, 2011). Furthermore, we saw the participants as a part of the interpretation process, allowing them to reflect on their statements deeply rather than dominating the interpretation by ourselves. For instance, the participants were able to express changes at any given point. This allowed
us to develop a precise representation of what has been said and enabled a dialogue consensus between participants and experimenters (Scheele and Groeben, 2020). We further discussed the findings in our expert research group under constant consideration of related work to achieve construct validity by theoretical matching. Lastly, we ensured data reliability through two timely separated coding processes to conduct consistency and stability (Mayring, 2014).

### 3.3 Data processing

Data processing has been highly structuralized using MAXQDA2020 and Microsoft Excel. The transcribed interviews and the allocation represented the primary data while we also considered the memos and field notes. We used axial coding to map relations between codes. Therefore, we performed a five-step data processing (see Figure 2). To gain a deep understanding of the data while staying open-minded, we performed (1) two loops of inductive coding of the transcripts, before we (2) developed a code system that was aggregated afterwards to (3) apply it to the index cards. This served as a preliminary step to (4) sort the coded allocations to trust and distrust constructs. In the last step (5), we coded the transcripts deductively using the step two aggregated code system.

![Figure 2. Structured data processing.](image)

### 4 Preliminary Results

A fundamental first finding is that participants have a linguistic tendency to talk about having trust or no trust rather than mentioning distrust. To express negative feelings, they used phrases like "it would naturally break the trust" (P4). Only one participant explicitly talked about experiencing "great distrust" (P1). This conflicts with the index cards' actual allocation, as participants commonly assigned cards to the concept of distrust, indicating that they did not distinguish both concepts explicitly. Also, participants often did not clearly separate the antecedents of trust and distrust, as P7 states that the absence of trust related factors would lead to distrust. Especially the (non-)fulfillment of expectations in the interaction can trigger both positive and negative feelings thus leading to trust or distrust.

Interestingly, the only person mentioning distrust has negative attitudes towards technology indicating that dispositional factors are likely to facilitate feelings of distrust. This assumption is supported by P8 who stated that distrust would only arise by negative prejudices. Other external factors like missing experience with chatbots or a severe disease do not lead to distrust, but missing trust. We also noted that distrust is likely to arise when minimum requirements are not met (e.g., correct spelling), when users fear data misuse or when the chatbot shows inappropriate behavior (e.g., by displaying adverts).

Regarding the allocation and the participant's conclusive assessment, we identified three determining mechanisms. First, the quantity of allocations has no impact on the actual perception of trust and distrust, as participants tend to assign an individual weight to the importance of determinants. Even if there is a high quantity of allocations to distrust, the overall tendency to trust may still arise. Although this finding contradicts our assumption of a generally overweighting of distrust, we argue that an emotions' arousal may be more important as one participant stated: "Well, this huge column indicating low distrust causes a little discomfort, because actually I didn't feel such a negative vibe so to speak,
such a negative feeling" (P4). This provides evidence that distrust only arises when a person actively experiences strong negative emotions. Thus, distrust is not triggered by the mere absence of trust.

Second, the constructs trust and distrust are perceived as context sensitive. Participants stressed that healthcare is a rather sensitive area, "such a topic, which one takes seriously" (P6). That is why there is more critical thinking involved in the process. For instance, one of the participants explained her allocation of "fraud" to high distrust with: "Yes, because it is something important, it is about health" (P3). In the specific context of a medical consultation, the fear of data misusage, dubious purposes or inconsistency may evoke strong feelings of distrust as patients expect care and benevolence.

Third, the analysis revealed that (dis-)trusting chatbots differs from interpersonal (dis-)trusting. For instance, one participant indicated that "machines tend to make even fewer mistakes" (P4) than human beings. Nevertheless, participants tend to trust a human being more because of general feelings, "simply because of the intuition coming from being a human who makes one believe that there is another person, which I trust more than a machine" (P5). We also noticed that some participants neutralize distrust towards the chatbot by rational reasoning (P1, P4, P5, P6). Correspondingly, several participants indicated that high transparency and the display of statistical figures on the data would enhance trust since both factors provide evidence for the chatbot's accuracy and reliability.

Lastly, we identified the need to differentiate between processual (dis-)trust and conclusive (dis-)trust. During trust- and distrust-building processes, the constructs can be negated, as well as co-exist. For instance, one of the participants stated that "distrust was something that arose in the process but then has dissolved" (P2). As we argued before, conclusive (dis-)trust may be a function of the overall experience that results from the intensity of positive and negative emotions.

5 Discussion

Our research aims to understand the differences between trust- and distrust-building processes towards diagnosis chatbots. Although it is hard to separate both concepts in interviews, we found evidence that some antecedents work on a single continuum while others only appear in the context of either trust or distrust. For instance, our results indicate that distrust only occurs when negative emotions are explicitly triggered by dispositional factors or the belief that the chatbot could potentially cause harm.

In the specific context of health chatbots, distrust-related factors entail negative beliefs concerning the system's purpose as patients are highly vulnerable and share sensitive personal information. Thus, we argue that antecedents like data privacy or non-profit purposes may be seen as self-evident hygiene factors whose fulfillment does not enhance trust, as only their absence causes distrust. In contrast, we state that the absence of trust-enhancing factors like providing transparency may only inhibit trust-building rather than causing distrust. With these first findings, we extend existing research by providing evidence that distrust towards chatbots in highly sensitive domains arises affectively while trust arises cognitively. This somehow contradicts research on trust in doctor-patient-relationships that emphasizes the importance of emotional trust (Thom and Campbell, 1997). Although users may build up an emotional attachment to intelligent agents (Glikson and Woolley, 2020), we barely found evidence for the trust-enhancing potential of affect-based components. One possible explanation is that affect-based trust mainly arises in reciprocal trusting relationships (McAllister, 1995). If the social component is missing, affective states may be more likely to arise negatively. Subjects then tend to apply coping mechanisms like rational reasoning and cognitive processing. However, this was only the case when perceived distrust was low, emphasizing the harmfulness of strong negative emotions.

Our research further supports that linguistic correlates of trust and distrust may differ as participants in our study mostly mentioned the absence of trust instead of using the word distrust. We noticed that it was hard for our participants to distinguish both concepts unambiguously as they seemed to have a binary way of thinking induced by the opposing use of trust and distrust in everyday language. However, it should be mentioned that our study was conducted in German language where trust (Vertrauen) and distrust (Mistrauen) share a common word root as they also do in English. Results from similar studies conducted in languages that clearly distinguish both concepts may be different and insightful (Van de Walle and Six, 2014).
Contrary to our predictions, we could not find evidence for a general overweighting of distrust yet. A possible explanation is that the experienced emotion's intensity may be more important than the mere sum of all emotions. This corresponds with the Peak-End-Rule assuming that people judge an experience by the most intense points and its end (Fredrickson and Kahneman, 1993). Considering the negativity bias, intense feelings of distrust may be more destructive than positive feelings of trust (McKnight and Chervany, 2001a). Thus, single negative moments with high arousal may lead to more harm than several moments of positive emotions could compensate. It is also to note that negative emotional states are further triggered by the context-sensitivity, which could be even higher for subjects with negative attitudes towards health chatbots. As distrust is more likely to prevail in high-risk situations (McKnight et al., 2004), we state that avoiding distrust reinforcing factors is crucial.

6 Limitations, Implications and Future Research

The main limitation of our study lies in its laboratory setting and the small sample size. To enhance reliability and validity, future studies should be carried out in real situations with a larger sample of actual patients to achieve more representative results. However, our pilot study's focus is to create an awareness for the high potential of considering trust and distrust distinctively and to open the discussion in this new research area. Thus, our results should be understood as a first profound reflection on (dis-)trusting processes towards health chatbots and as a starting point for future research.

Since it is difficult to separate the linguistically closely related concepts of trust and distrust in interviews (Gefen et al., 2020), we aim to conduct an additional study considering the users' tendency to trust or distrust diagnosis chatbots. It may be promising to conduct focus groups with either distrusting or trusting subjects to identify possible structural differences regarding their reasons for (dis-)trusting. Also, we will capture a person's disposition to (dis-)trust health chatbots since skeptical people may be more sensitive to distrust and negative emotions. Based on our results and prior studies (Lee et al., 2015), we will further conceptualize trust cognitively and distrust affectively in subsequent quantitative studies to validate our first findings. In addition, it could be insightful to include physiological measurements or facial expressions to capture the valence and arousal of experienced emotions during the interaction with a chatbot to study (dis-)trust-building processes deeply.

In the specific context of a medical consultation, patients wish for emotional trust that is less likely to arise in interactions with inanimate entities due to the missing reciprocal relationship (McAllister, 1995). Thus, it is not surprising that subjects experience higher levels of trust (lower levels of distrust) towards physicians. Although we barely found evidence for a trust-enhancing effect of emotion-related factors, future studies should investigate the drivers of positive affections towards chatbots. Experiencing positive emotions may diminish distrust-building towards health chatbots as it is mainly driven affect-based. In this sense, one interesting research area is the investigation of anthropomorphic design cues on (dis-)trust. Although social characteristics of virtual agents may have positive effects under specific circumstances (Blut et al., 2021), too much human-likeness may trigger distrust-related emotions like suspicion or doubt (Glikson and Woolley, 2020; Stein and Ohler, 2017). Therefore, it is vital to investigate in how far users expect emotion-related trust from health chatbots.

Conclusively, we encourage researchers to focus on distrust avoidance in IS research, especially in sensitive environments. We argue that a deep understanding of the arising of distrust towards chatbots should gain more theoretical and empirical interest. Much research postulates that trust enhances the willingness to adopt new technology, but the absence of distrust may be a prerequisite for it.

Despite our recommendation to understand and reduce drivers of distrust towards diagnosis chatbots, there are some threats to consider. First, a certain degree of distrust towards novel technologies has a protective function. Therefore, the aim should be to create appropriate rather than maximum trust (Lee and See, 2004). Second, the use of health chatbots may endanger the intimate doctor-patient-relationship (LaRosa and Danks, 2018). From an ethical point of view, developers and researchers should always respect the uniqueness of interpersonal relationships. Therefore, diagnosis chatbots should rather be seen as a supplement, not as a substitute (Powell, 2019).
References


