



The Trustworthiness of Voice Assistants in the Context of Healthcare Investigating the Effect of Perceived Expertise on the Trustworthiness of Voice Assistants, Providers, Data Receivers, and Automatic Speech Recognition

Carolin Wienrich*, Clemens Reitelbach and Astrid Carolus

Julius Maximilian University of Würzburg, Würzburg, Germany

OPEN ACCESS

Edited by:

Stefan Hillmann,
Technical University Berlin, Germany

Reviewed by:

Christos Troussas,
University of West Attica, Greece
Rubén San-Segundo,
Polytechnic University of Madrid,
Spain

*Correspondence:

Carolin Wienrich
carolin.wienrich@uni-wuerzburg.de

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Computer Science

Received: 24 March 2021

Accepted: 19 May 2021

Published: 17 June 2021

Citation:

Wienrich C, Reitelbach C and
Carolus A (2021) The Trustworthiness
of Voice Assistants in the Context of
Healthcare Investigating the Effect
of Perceived Expertise on the
Trustworthiness of Voice Assistants,
Providers, Data Receivers, and
Automatic Speech Recognition.
Front. Comput. Sci. 3:685250.
doi: 10.3389/fcomp.2021.685250

As an emerging market for voice assistants (VA), the healthcare sector imposes increasing requirements on the users' trust in the technological system. To encourage patients to reveal sensitive data requires patients to trust in the technological counterpart. In an experimental laboratory study, participants were presented a VA, which was introduced as either a "specialist" or a "generalist" tool for sexual health. In both conditions, the VA asked the exact same health-related questions. Afterwards, participants assessed the trustworthiness of the tool and further source layers (provider, platform provider, automatic speech recognition in general, data receiver) and reported individual characteristics (disposition to trust and disclose sexual information). Results revealed that perceiving the VA as a specialist resulted in higher trustworthiness of the VA and of the provider, the platform provider and automatic speech recognition in general. Furthermore, the provider's trustworthiness affected the perceived trustworthiness of the VA. Presenting both a theoretical line of reasoning and empirical data, the study points out the importance of the users' perspective on the assistant. In sum, this paper argues for further analyses of trustworthiness in voice-based systems and its effects on the usage behavior as well as the impact on responsible design of future technology.

Keywords: voice assistant, trustworthiness, trust, anamnesis tool, expertise framing (Min5-Max 8)

INTRODUCTION

Voice-based artificial intelligence systems serving as digital assistants have evolved dramatically within the last few years. Today, Amazon Echo or Google Home is the most popular representatives of the fastest-growing consumer technology (Hernandez, 2021; Meticulous Market Research, 2021). On the one hand, voice assistants (VAs) engage human users in direct conversation through a natural language interface leading to promising applications for the healthcare sector, such as diagnosis and therapy. On the other hand, their constituting features to recognize, process, and produce human language results in this technology to resemble human-human interaction. Attributing some kind of

humanness to technology arouses (implicit) assumptions about the technological devices and affects the user's perception and operation of the device. The media equation approach postulates that the social rules and dynamics guiding human-human interaction similarly apply to human-computer interaction (Reeves and Nass, 1996). Using voice assistants in official application areas involving sensitive data such as medical diagnoses draws attention to the concept of trust: if patients were to reveal personal, sensitive information to the voice-based systems, they would need to trust them. Consequently, questions of the systems' trustworthiness arise asking for features of voice assistants, which might affect the patients' willingness to trust them in a medical context. Results stemming from studies investigating trust in human-human interactions revealed that ascribed expertise is a crucial cue of trust (Cacioppo and Petty, 1986; Chaiken, 1987; Chaiken and Maheswaran, 1994). Reeves and Nass (1996) transferred the analysis of expertise and trust to human-technology interactions. They showed that designating devices (here: a television program) as "specialized" results in more positive evaluations of the content they presented. Many other studies replicated their approach and framed a technological device or a technological agent as a specialist showing that users ascribed a certain level of expertise and evaluated it (implicitly) as more trustworthy (Koh and Sundar, 2010; Kim, 2014, Kim, 2016; Liew and Tan, 2018).

Voice assistants gain in importance in healthcare contexts offering promising contributions in the area of medical diagnosis, for instance. However, both the analysis and the understanding of the psychological processes characterizing the patient-voice assistant interaction are still in their early stages. Similarly, the effects of the assistant's design on the perception of expertise and the evaluation of trust are still in their infancy. Thus, the present paper addressed the following research question: How does framing a voice assistant as a specialist affect the user's perception of its expertise and its trustworthiness?

To gain first insights into the process of patients' perception of expertise of voice-based systems and their willingness to trust in them, a laboratory study was conducted in which participants interacted with a voice assistant. The assistant was introduced as a diagnostic tool for sexual health, which asked a list of questions about sexual behavior, sexual health, and sexual orientation to determine the diagnosis. In a first step, and in accordance with the approach of Reeves and Nass (1996), we manipulated the level of expertise of the voice assistant, which introduced itself as either a "specialist" or a "generalist". In line with established approaches investigating the trustworthiness of technology (e.g., McKnight et al., 1998; Söllner et al., 2012), we compared the participants' perceived trustworthiness of the "specialist" vs. the "generalist" VA. Additionally, we compared the assessments of further source layers of trustworthiness, namely of the platform provider, the provider of the tool, the data receiver, and of automatic speech recognition in general. Moreover, to account for additional explanatory value of interindividual differences in the trustworthiness ratings, we asked for participants' dispositions and characteristics such as their disposition to trust and their tendency to disclose sexual information about themselves. Finally, we analyzed the different source layers of

trustworthiness to predict the trustworthiness of the VA based on the trustworthiness of the other source layers. In sum, the present paper showed for the first time that a short written introduction and a "spoken" introduction presented by the VA itself were sufficient to affect the users' perception and their trust in the system significantly. Hence it addresses a human-centered approach to voice assistants to show that small design decisions determine user's trust in VA in a safety-critical application field.

RELATED WORK

Voice Assistants in Healthcare

While voice-based artificial intelligence systems have increased in popularity over the last years, their spectrum of functions, their field of applications, and their technological sophistication have not been fully revealed but are still in their early stages. Today's most popular systems—Amazon Echo (AI technology: Amazon Alexa) or Google Home (AI technology: Google Assistant)—presage a variety of potential usage scenarios. However, according to usage statistics, in a private environment, voice assistants are predominantly used for relatively trivial activities such as collecting information, listening to music, or sending messages or calls (idealo, 2020). Beyond private usage scenarios, voice assistants are applied in professional environments such as industrial production or technical service (e.g., Baumeister et al., 2019), voice marketing, or internal process optimization (Hörner, 2019). In particular, the healthcare sector has been referred to as an emerging market for voice-based technology. More and more use cases emerge in the context of medicine, diagnosis, and therapy (The Medical Futurist, 2020) with voice assistants offering promising features in the area of anamnesis. Particularly the possibility to assess data remotely gains in importance these days. Recently, chatbots were employed to collect the patients' data, their medical conditions, their symptoms, or a disease process (ePharmaINSIDER, 2018; The Medical Futurist, 2020). While some products provide only information (e.g., OneRemission), others track health data (e.g., Babylon Health) or check symptoms and make a diagnosis (e.g., Infermedica). Until today, only a few solutions have integrated speech recognition or direct connection to VA, such as Alexa *via* skills (e.g., Sensely, Ada Health, GYANT). The German company *ignimed UG* (<https://ignimed.de/>) takes a similar approach: based on artificial intelligence, the patient's information is collected and transmitted to the attending physician, who can work with the patient. Although these voice assistants are used for similar purposes, which all require user to trust the system, users' perceived trustworthiness of voice assistants in healthcare has not been investigated yet.

The medical context imposes different requirements on the system than private usage scenarios do. Data revealed here are more personal and more sensitive, resulting in increasing requirements regarding the system's security and trustworthiness. Consequently, besides focusing on technological improvements of the system, its security or corresponding algorithms, research needs to focus on the patients' perception of the system and their willingness to interact with them in a health-related context. Exceeding the question of which gestalt design impacts both usability and user

experience, the field of human-computer interaction needs to ask for features affecting the patients' perceived trustworthiness of the technological counterpart they interact with. One promising approach is analyzing and transferring findings from human-human interactions to human-computer or human-voice assistant interactions. Following the media equation approach of Nass and colleagues, this study postulates similarities between the human counterpart and the technological counterpart, which results in psychological research to be a fruitful source of knowledge and inspiration for the empirically based design of voice assistants in a medical context.

Interpersonal Trust: The Role of Expertise

Interpersonal interactions are characterized by uncertainty and risks since the behavior of the interaction partner is unpredictable—at least to a certain extent. Trust defines the intention to take the risks of interaction by reducing the perceived uncertainty and facilitating the willingness to interact with each other (Endreß, 2010). In communication contexts, trust refers to the listener's degree of confidence in, and level of acceptance of, the speaker and the message (Ohanian, 1990). Briefly spoken, trust in communication refers to the listener's trust in the speaker (Giffin, 1967). According to different models of trust, characteristics of both the *trustor* (the person who gives the trust) and the *trustee* (the person who receives the trust) determine the level of trust (e.g., Mayer et al., 1995). The dimensions of competence, benevolence, and integrity describe the trustee's main characteristics (see, for example, the meta-analysis of McKnight et al. 2002). The perceived trustworthiness of trustees increases with increasing perceived competence, benevolence, or integrity. In communication contexts, the term *source credibility* closely refers to the trustor's perceived trustworthiness in terms of the trustee. It refers to the speaker's positive characteristics that affect the listener's acceptance of a message. The *source-credibility model* and the *source-attractiveness model* concluded that three factors, namely, expertness, trustworthiness, and attractiveness, underscore the concept of source credibility (Hovland et al., 1953). In this context, expertise is also referred to as authoritativeness, competence, expertness, qualification, or being trained, informed, and educated (Ohanian, 1990). In experiments, the perceived expertise of speakers was manipulated by labeling them as “Dr.” vs. “Mr.” or as “specialist” vs. “generalist” (e.g., Crisci and Kassinove, 1973). The labels served as cues that can bias the perception of the competence, benevolence, or integrity of trustees or communicators and the perception of trust.

When comprehending the underlying effects of information processing, well-established models of persuasion reveal two routes of processing—the heuristic (peripheral) route and the systematic (central) route (e.g., the *heuristic-systematic model*, *HSM* by Chaiken (1987); the *elaboration likelihood model*, *ELM* by Cacioppo and Petty (1986). The heuristic (peripheral) route is based on judgment-relevant cues (e.g., source's expertise) and needs less cognitive ability and capacity than the systematic (central) route, which is based on judgment-relevant information (e.g., message content). Typically, individuals will

prefer the heuristic route as the more parsimonious route of processing if they trust the source, particularly if cues activate one of the three trustworthiness dimensions (Koh and Sundar, 2010). For example, individuals will perceive more trustworthiness when a person is labeled as a “specialist” compared to a “generalist” since a specialist sends more cues of expertise and activates the dimension of competence (Chaiken, 1987; Chaiken and Maheswaran, 1994). Remarkably, the effect will endure even if both the specialist and the generalist possess objectively the same level of competence or expertise. Consequently, individuals interacting with a specialist are more likely to engage in heuristic processing and implicitly trust the communicator (Koh and Sundar, 2010).

Regarding the resulting level of trust, the trustor's characteristics were found to moderate the impact of the trustee's characteristics. First, the perceived level of expertise depends on the interindividual differences in the processing of information. The outlined indicators of trust need to be noticed and correctly interpreted to have an effect. Furthermore, the individual's personality and experiences were shown to influence the perception of trustworthiness. Finally, an individual's disposition to trust as the propensity to trust other people has been shown to be a significant predictor (Mayer et al., 1995; McKnight et al., 2002).

To summarize, research in various areas revealed that perceived expertise affects the trustee's trustworthiness (e.g., commercial: Eisend, 2006, health: Gore and Madhavan, 1993; Kareklas et al., 2015), general review see Pornpitakpan (2004). This perception is also affected by the trustor's characteristics (e.g., the disposition to trust). With the digital revolution proceeding, technology has become increasingly interactive, assembling human-human interaction to an increasing extent. Today, an individual does not only interact with other human beings but also with technological devices. These new ways of human-technology interaction require the individual to trust in technological counterparts. Consequently, the question arises whether the outlined mechanisms of trust can be transferred to non-human technological counterparts.

Trust in Technology: The Role of Expertise

The media equation approach postulates that social rules and dynamics, which guide human-human interaction apply to human-computer interaction similarly (Reeves and Nass, 1996; Nass and Moon, 2000). To investigate the media equation assumptions, Nass and Moon (2000) established the CASA paradigm (i.e., computer as social actors) and adopted well-established approaches from research on human-human interaction to the analysis of human-computer interactions. Many experimental studies applying the CASA paradigm demonstrated that individuals tend to transfer social norms to computer agents, for example, gender and ethnic stereotypes and rules of politeness and reciprocity (Nass and Moon, 2000). More specific in the context of trust, experimental studies and imagine-based approaches revealed that trust-related situations activate the same brain regions regardless of whether the counterpart is a human being or a technological agent (Venkatraman et al., 2015; Riedl et al., 2013). Consequently, researchers concluded that there

are similar basic effects elicited by human and technological trustees (Bär, 2014).

However, when interacting with a non-human partner, the trustee's entity introduces several interwoven levels of trustworthiness, referred to as *source layers* in the following (Koh and Sundar, 2010). The trustors can trust the technical device or system (e.g., VA) itself. Moreover, they could also refer to the provider, the domain, or the human being "behind" the system, such as the person who receives the information (Hoff and Bashir, 2015). Similar to interpersonal trust, three dimensions determine the perceived trustworthiness of technology: performance (analogous to human competence), clarity (analogous to human benevolence), and transparency (analogous to human integrity) (Backhaus, 2017). As known from interpersonal trust, credibility factors bias the perception of the expertise ascribed to the technology. For example, in the study by Reeves and Nass (1996), participants watched and evaluated a *news* or a *comedy* television program. In the study, they were assigned to one of two conditions: the "specialist television" or the "generalist television". The conditions differed regarding the instruction presented by the experimenter, who referred to the television as either the "news TV" or the "entertainment TV" (specialist condition) or to "usual TV" (generalist condition). Findings indicated that individuals evaluated the content presented by the specialist TV set as more positive than the content of the generalist TV set—even though the content was completely identical. The results have been replicated in the context of specialist/generalist television channels (e.g., Leshner et al., 1998), smartphones (Kim, 2014), or embodied avatars (Liew and Tan, 2018). Additionally, in the context of e-health, the level of expertise was shown to affect the perception of trustworthiness (e.g., Bates et al., 2006). Koh and Sundar (2010) explored the psychological effects of expertise (here: specialization) of web-based media technology in the context of e-commerce. They distinguished multiple indicators or sources of trustworthiness (i.e., computer, website, web agent), they referred to as "source layers of web-based mass communication". In their study, they analyzed the effects on individuals' perceptions of expertise and trust distinguishing between these source layers. In their experiment, participants interacted with media technology (i.e., computer website, web agent), which was either labeled as specialist ("wine computer", "wine shop" or "wine agent") and generalist ("computer", "e-shop" or "e agent"). Again, only the label but not the content differed between the two experimental conditions. Findings supported the positive effects of the specialization label. Participants reported greater levels of trust in specialist media technology compared to generalist media technology with the "specialized" web agent eliciting the strongest effects (compared to "specialized" website or computer). Consequently, this study focusses on the multiple indicators or the multiple source layers contributing to the trustworthiness of a complex technological system. According to which source layer is manipulated the users' assessment of trustworthiness might differ fundamentally.

To summarize, research so far focused mostly on the credibility of online sources (e.g., websites), neglecting other

technological agents like voice-based agents, which currently capture the market in the form of voice bots, voice virtual assistants or smart speaker skills. Furthermore, research so far focused on the engineering progress resulting in increasingly improved performances of the systems but tends to neglect the human user, who will interact with the system. As outlined above, in usage scenarios involving sensitive data, the human users' trust in the technological system is a fundamental requirement and a necessary condition of the user opening up to the system. Voice-based systems in a healthcare context need to be perceived as trustworthy agents to get a patient to disclose personal information. However, scientific studies so far reveal a lack of detailed and psychologically arguing analyses and empirical studies investigating the perceived trustworthiness of voice assistants. The present study aims for first insights into the users' perception of the trustworthiness of voice assistants in the context of healthcare raising the following research questions: 1) Does the introduction of a voice assistant as an expert increase its trustworthiness in the context of healthcare? 2) Do the users' individual dispositions influence the perceived trustworthiness of the assistant? 3) How do the levels of perceived trustworthiness of the multiple source layers (e.g., assistant tool, provider, data receiver) interact with each other?

Outline of the Present Study

To answer the research questions, a laboratory study was conducted. Participants interacted with the *Amazon Echo Dot*, Amazon's voice assistant referred to as "the tool" in the following. The VA was introduced to the participants as an "anamnesis tool for sexual health and disorders", which would ask questions about the participants' sexual behavior, their sexual health, and their sexual orientation. Following the approach of Nass and Reeves (1996), participants were randomly assigned to one out of two groups, which differed by one single aspect: the labeling of the VA. Participants received a written instruction in which the VA was either referred to as a "specialist" or a "generalist". Furthermore, at the beginning of the interaction, the VA introduced itself as either a "specialist" or a "generalist". In line with studies investigating trust in artificial agents (e.g., McKnight et al., 1998; Söllner et al., 2012) and studies including multiple sources layers of trustworthiness (e.g., Koh and Sundar, 2010), we distinguished between different source layers of perceived trustworthiness of our setting: the perceived trustworthiness of the VA tool itself, the provider of the tool (i.e., a German company), the platform provider (i.e., Amazon), automatic speech recognition in general, and the receiver of the data (i.e., the attending physician). Furthermore, participant's individual characteristics, i.e., the disposition to trust and the tendency to sexual self-disclosure, were considered.

METHOD

Participants

The 40 participants (28 females, 12 males; average age = 22.45 years; $SD = 3.33$) were recruited via personal contact or the university recruitment system offering course credit. All

participants were German native speakers. Except for one, all participants were students. 80% of them reported having already interacted with a voice assistant. However, when analyzing the duration of these interactions, the sample's experience was rather limited: 75% reported to have interacted with a VA for less than 10 h and 45% for less than 2 h in total.

Task, Manipulation, Pre-test of Manipulation and Pre-test of Required Trust

During the experiment, participants interacted with a VA, *Amazon Echo Dot* (3rd Generation, black). While the VA asked them questions about their sexual behavior, sexual health, and sexual orientation, participants were instructed to answer these questions as honestly as possible using speech input. Participants were randomly assigned to one of two groups ($n = 20$ per group), which only differed regarding the label of the VA. In an introduction text, the VA was introduced as an anamnesis tool, labeled either a "specialist" (using words such as "specialist," "expert") or as a "generalist" (e.g., "usual," "common"). Additionally, the VA introduced itself in two ways. In the "specialist" condition, it referred to itself as a "special tool for sexual anamnesis" and in the general condition as a "general survey tool."

A pre-study ensured the effect of this manipulation and trust to be a prerequisite of answering the anamnesis questions. In an online survey, 30 participants read one of the two introduction texts and described the tool in their own words, afterward. A content analysis of their descriptions showed that participants followed the labeling of the text using compatible keywords to describe the device ("specialist" condition: e.g., special, expert; vs. "generalist" condition: normal, common). However, because only twelve participants used at least one predefined condition-related keyword, the experimental manipulation was strengthened by adding more keywords to the instruction text. The final manipulation text is attached to the additional material. Since the VA's perceived trustworthiness was the main dependent variable, the second part of the pre-test ensured that answering the sexual health-related questions required trust. All anamnesis questions were presented to the participants, who rated how likely they would answer each question. The scale ranged from 100 (very likely) to 0 (no, too private) with lower scores indicating higher levels of required trust to answer the question. Questions were clustered in four categories: puberty, sexual orientation, diseases/hygiene and sexual activity. Results showed that questions regarding puberty (average rating = 75.78) and sexual orientation (67.01) required less trust than diseases/hygiene (56.47) and sexual activity (50.75). To ensure a minimum standard of required trust, one question of the puberty category was removed. Furthermore, four conditional questions were added to the categories of diseases/hygiene and sexual activity, which would ask for more detailed information if previous questions were answered with "yes".

To assess the perceived trustworthiness of the tool, different source layers were considered. First, the trustworthiness of the tool provider, German company, *ignimed UG*, had to be evaluated. Second, since the VA tool was connected to

Amazon Echo Dot, the trustworthiness of the platform provider (Amazon) was assessed. Third, the trustworthiness of the potential perceiver to the data (gynecologists/urologist) was rated. Finally, we added automatic speech recognition as a proxy for the underlining technology, which the participant also rated in terms of trustworthiness. Note, the experimental manipulation of expertise referred only to the tool itself. Consequently, the VA tool represents the primary source layer, while others refer to further source layers.

The Sexual Health Anamnesis Tool: Questions the VA Asked

After introducing itself, the VA started the anamnesis conversation, which involves 21 questions (e.g., *Do you have venereal diseases?—Which one?*). Four categories of questions were presented: puberty (e.g., *What have been the first signs of your puberty?*), diseases/hygiene (e.g., *Have you ever had one or more sexual diseases?*), sexual orientation (e.g., *What genders do you have sexual intercourse with?*) and sexual activity of the past 4 weeks (e.g., *How often have you had sexual intercourse in the past 4 weeks?*). The complete list of final measurements follows below.

Measurements

After finishing the conversation with the VA, participants answered a questionnaire presented via LimeSurvey on a 15.6" laptop with an attached mouse. The measures of the questionnaire are presented below:

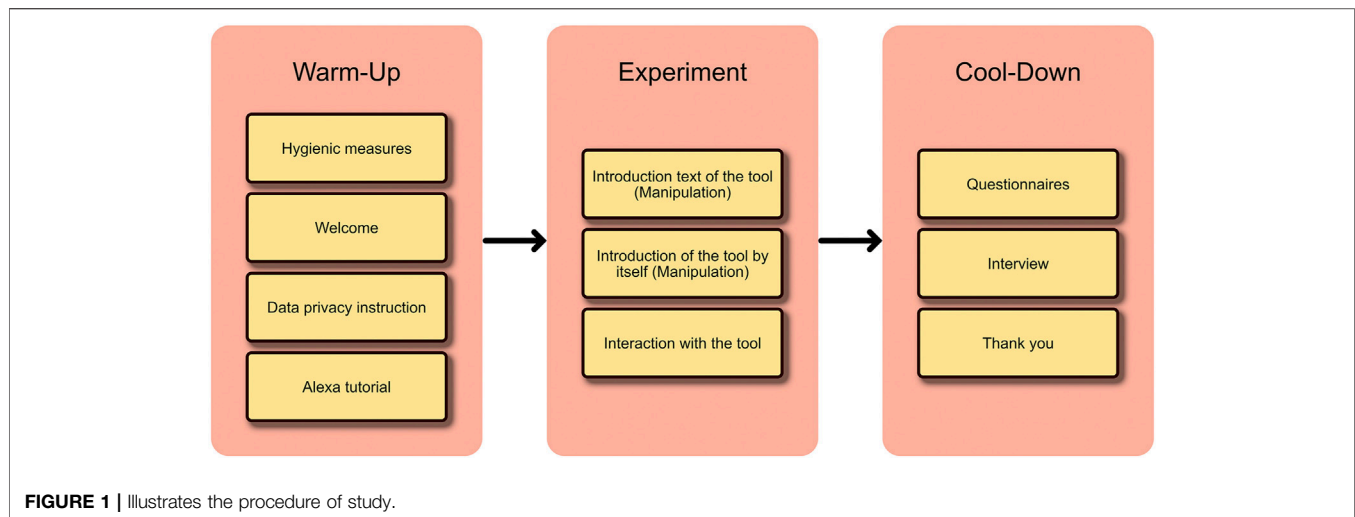
Perceived Trustworthiness of Source Layers

To measure the **trustworthiness of the VA** three questions adapted from Corritore et al. (2003) were asked (e.g., *I think the tool is trustworthy*). Additionally, an overall item adapted from Casaló et al. (2007) was presented (*Overall I think that the tool is a safe place for sensitive information*). All questions were answered on a 7-point Likert scale, ranging from *not true at all* to *very true*.

Questions concerning institutional trust from the SCOUT Questionnaire (Bär et al., 2011) were transferred to assess the perceived trustworthiness of the tool provider, the platform provider, and automatic speech recognition. Items were answered on a 5-point Likert scale, ranging from *not agree at all* to *agree totally*. Five questions assessed the **tool provider's perceived trustworthiness** (e.g., *I believe in the honesty of the provider*) and the **platform provider** (same five questions). Four questions assessed the **trustworthiness of automatic speech recognition** (*Automatic speech recognition is trustworthy technology*). Finally, the **data receiver's perceived trustworthiness**, namely, the participant's gynecologist/urologist, the KUSIV3-questionnaire, was used (Beierlein et al., 2012). It includes three questions (e.g., *I am convinced that my gynecologist/urologist has good aims*) on a 5-point Likert scale, ranging from *not agree at all* to *agree totally*.

Individual Characteristics

The disposition of trust was measured with three statements (e.g., *For me, it is easy to trust persons or things*), assessed on a 5-point Likert scale (ranging from *not agree at all* to *agree totally*) taken



from the SCOUT Questionnaire (Bär et al., 2011). The tendency to sexual self-disclosure was measured with the sexual self-disclosure scale (Clark, 1987). Participants rated four questions, two on a 5-point Likert scale (e.g., *How often do you talk about sexuality?*, ranging from *never-rarely* to *very often*) and two using a given set of answers (e.g., *With whom do you talk about sexuality?* *mother* | *father* | *siblings* | *partner* | *friends(male)* | *friends(female)* | *doctors* | *nobody* | *other*).

Manipulation Check

To measure how strong the participants believe that the tool is a "specialist" or "generalist," two questions (e.g., *The survey tool has high expertise in the topic*) were answered using a 5-point Likert scale.

Procedure

The study took about 40 min, starting with COVID-19 hygienic routines (warm-up phase: washing and antisepticizing hands, answering a questionnaire and wearing a mouth-nose-mask). Since the experimental supervisor left the room for the actual experiment, participants could discard their face masks during the interaction with the VA. In the warm-up phase, participants were instructed to do a short tutorial with the VA, which asked some trivial questions (e.g., *How is the weather?* or *Do you like chocolate?*). When the participants confirmed to be ready to start the experiment, they were instructed to read the introduction text about the anamnesis tool and to start the interaction with the VA (experimental phase). After finishing, the participants answered the questionnaires and were briefly interviewed about the experience with the VA by the supervisor (cool-down phase). **Figure 1** illustrates the procedure of the study and the three experimental phases.

Design, Hypothesis

Accordingly to previous studies (e.g., Gore and Madhavan, 1993; Reeves and Nass, 1996; Kareklas et al., 2015), the experiment followed a between-subjects design with two conditions ("specialist" or "generalist" VA). In line with the first research question, referring to the effects of perceived expertise on perceived trustworthiness (differentiated regarding source layers), the first hypotheses postulated that the perceived

trustworthiness (across the source layers) would be higher in the specialist condition than in the generalist condition. The second research question addressed the impact of individual dispositions on perceived trustworthiness. In line with Mayer et al. (1995) and McKnight et al. (2002), the second hypotheses assumed that higher individual trust-related dispositions result in higher trustworthiness ratings. Finally, the third research question explorative asked whether the perceived trustworthiness of the multiple source layers (e.g., the assistant tool, the providers, the receiver) interact with each other (Koh and Sundar (2010)). **Table 1** gives an overview of the hypothesis.

Data Analyses

Data have been prepared as proposed by the corresponding references. To facilitate the comparability of measures of trustworthiness, items answered on a 7-point Likert scale were converted to a 5-point scale. Five t-tests for independent groups (specialist condition vs. generalist condition) were conducted to test the first hypotheses. The second group of hypotheses was tested, conducting five linear regression analyses with the source layers' trustworthiness as the five criteria variables and the individual characteristics as the predictor variables. Finally, a linear regression analysis was conducted regarding the explorative research question with trustworthiness of the VA as the criteria and the other layers of trustworthiness as the predictors. The following section report means (M), standard deviations (SD) of scales as well as the test statistic parameters such as the *t*-value (*t*), and *p*-value (*p*).

RESULTS

Impact of Expert Condition on Source Indicators' Perceived Trustworthiness

As expected, the perceived trustworthiness of the tool was higher in the specialist condition ($M = 3.206$, $SD = 1.037$) than in the generalist condition ($M = 2.634$, $SD = 0.756$). However, the result was just not significant ($t_{(38)} = 2.019$, $p = 0.051$, $d = 0.638$). Contrary to our expectations, the perceived trustworthiness of the platform provider, the tool provider, the data receiver, and the automatic speech recognition did not differ significantly between

TABLE 1 | Hypotheses and a short overview of corresponding results.

Class of hypotheses	Dependent variables	Results
H1: Source indicators' perceived trustworthiness is higher in the specialist condition than in the generalist condition.	H1a: tool's trustworthiness H1b: Tool provider's trustworthiness H1c: Platform provider's trustworthiness H1d: Data receiver's trustworthiness H1e: Trustworthiness of automatic speech recognition	Confirmed with adjusted analysis. Confirmed with adjusted analysis. Confirmed with adjusted analysis. Confirmed with adjusted analysis. Confirmed with adjusted analysis.
H2: Higher individual trust-related dispositions and tendencies result in higher trustworthiness ratings.	H2a: disposition to trust H2b: Tendency to sexual self-disclosure	Confirmed for the trustworthiness of automatic speech recognition. Confirmed for the tool's trustworthiness. Confirmed by trend for the trustworthiness of automatic speech recognition.
RQ3: Does the perceived trustworthiness for multiple source indicators interact with each other?	-	As higher the tool provider's scores, as higher is the perceived trustworthiness of the tool itself. By trend: As higher the speech recognition scores, as higher is the tool's perceived trustworthiness. By trend: As higher the platform provider's (i.e., Amazon) scores, the tool's perceived trustworthiness is lower.

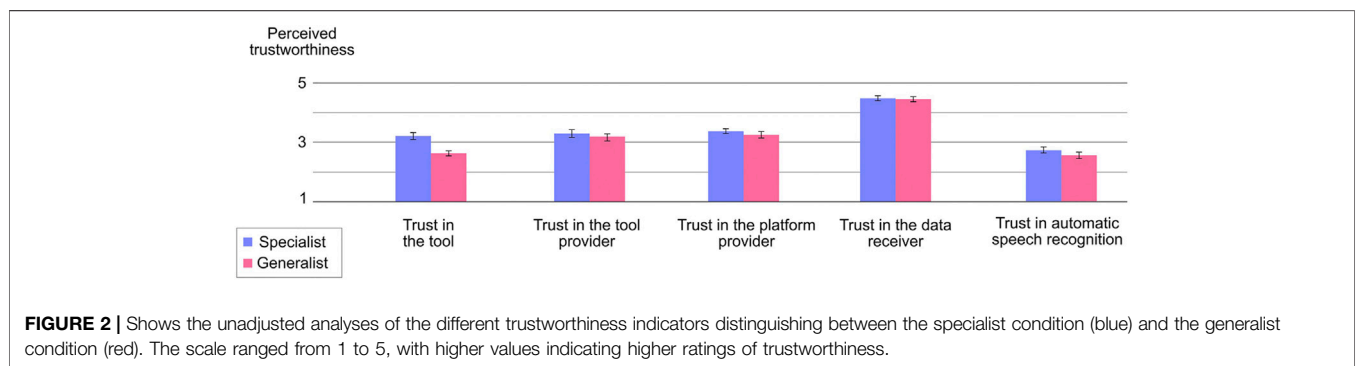


FIGURE 2 | Shows the unadjusted analyses of the different trustworthiness indicators distinguishing between the specialist condition (blue) and the generalist condition (red). The scale ranged from 1 to 5, with higher values indicating higher ratings of trustworthiness.

TABLE 2 | Results of unadjusted analyses.

Indicator of trustworthiness	Specialist		Generalist		Significance test
	M	SD	M	SD	
Tool	3.206	1.037	2.634	0.756	$t(38) = 2.019, p = 0.051, d = 0.638$
Tool provider (German company)	3.290	1.145	3.190	1.053	$t(38) = 0.287, p = 0.775, d = 0.091$
Platform provider (amazon)	3.250	0.969	3.370	0.700	$t(38) = -0.449, p = 0.656, d = 0.142$
Data receiver (gynecologists/urologist)	4.483	0.791	4.450	0.767	$t(38) = 0.135, p = 0.893, d = 0.042$
Automatic speech recognition	2.725	0.862	2.550	0.955	$t(38) = 0.608, p = 0.547, d = 0.192$

the conditions (see **Figure 2** and **Table 2** for descriptive results and t-test results of unadjusted analyses).

Impact of Manipulation Check Success on Source Indicators' Perceived Trustworthiness

Although a pre-test confirmed the manipulation of the two conditions, the manipulation check of the main study revealed a lack of effectivity: two control questions showed that the specialist

tool was not rated as significantly “more special” than the generalist tool (see **Table 3** for descriptive and t-test results). Consequently, the assignment to the two groups did not result in significantly different levels of perceived expertise of the VA.

Consequently, we needed to adjust the statistical analyses of the group comparisons. We re-analyzed the ratings of the control questions (asking for the VA's expertise). Based on the actually perceived expertise, participants were divided into two groups with ratings below and above the averaged scales' median ($MD = 3.333$). Independently of the intended manipulation, 22

TABLE 3 | Results of manipulation check.

Control question: "The survey tool ...	Specialist M (SD)	Generalist M (SD)	Significance test
... has high expertise in the topic."	3.300 (0.923)	3.050 (0.887)	$t(df) = 0.873, p = 0.388, d = 0.276$
... was developed for a special purpose."	3.650 (1.268)	3.350 (1.040)	$t(df) = 0.818, p = 0.419, d = 0.259$

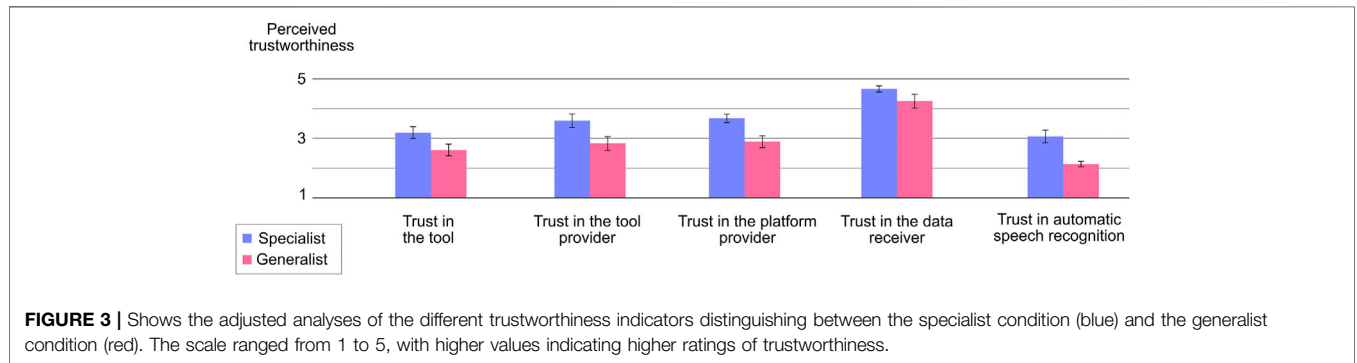


FIGURE 3 | Shows the adjusted analyses of the different trustworthiness indicators distinguishing between the specialist condition (blue) and the generalist condition (red). The scale ranged from 1 to 5, with higher values indicating higher ratings of trustworthiness.

participants revealed higher ratings of the VA's expertise (> 3.5; group 1) indicating that they rather perceived a specialist tool. In contrast, 18 participants revealed lower ratings of the expertise (< 3.0; group 2) indicating that they rather perceived a generalist tool. In sum, the re-analyses (referred to as the "adjusted analysis" below) will analyze the perceived trustworthiness of participants, who actually perceived the VA as a specialist or a generalist independently of the intended manipulation (see **Figure 3**).

In line with the hypotheses, the **perceived trustworthiness of the tool** would be significantly higher, if the tool was actually perceived as a specialist ($M = 3.182, SD = 0.938$) compared to the perception of a generalist ($M = 2.599, SD = 0.836; t_{(38)} = 2.051, p = 0.045, d = 0.652$). Also in line with expectations, the **platform provider's perceived trustworthiness** would be significantly higher, if the VA was perceived as a specialist ($M = 3.581, SD = 1.313$) than a generalist ($M = 2.822, SD = 1.170, t_{(38)} = 2.318, p = 0.025, d = 0.737$). The same applies to the **provider's perceived trustworthiness**: the provider of the specialist tool was rated to be significantly more trustworthy ($M = 3.663, SD = 0.066$) than the provider of a generalist tool ($M = 2.878, SD = 0.984; t_{(38)} = 3.331, p = 0.02, d = 1.503$). Likewise, the **perceived trustworthiness of automatic speech recognition** was significantly higher for a perceived specialist ($M = 3.057, SD = 0.994$) than a generalist ($M = 2.125, SD = 0.376; t_{(38)} = 3.757, p = 0.01, d = 1.194$). However, the data receiver's perceived trustworthiness did not differ significantly between a specialist ($M = 4.652, SD = 0.488$) and a generalist ($M = 4.241, SD = 0.982; t_{(38)} = 1.722, p = 0.093, d = 0.547$).

Additional Predictors of the Perceived Trustworthiness

Individual Characteristics

Linear regression analyses were conducted with the five trustworthiness source layers as the criteria variables and the general disposition to trust and to disclose sexual health information as predictor variables. Only two regressions

involved significant predictions: the perceived trustworthiness of the VA and of the speech recognition in general. First, the prediction of the tool's perceived trustworthiness was significant ($\beta = 0.655, t_{(37)} = 2.056, p = 0.047$) with the tendency to disclose health information contributing significantly: the higher this tendency, the higher the perceived trustworthiness of the tool. Second, the prediction of the automatic speech recognition's trustworthiness was significant, with the participants' disposition to trust contributing significantly to the prediction ($\beta = 0.306, t_{(37)} = 2.276, p = 0.029$). Moreover, the tendency to disclose sexual health information contributed substantially but only by trend ($\beta = 0.370, t_{(37)} = 1.727, p = 0.093$).

Further Source Layers of Trustworthiness

The final regression analysis investigated whether the further source layers of trustworthiness indicators (providers, speech recognition, receiver) predicted the tool's trustworthiness. Results revealed that the **trustworthiness of the provider** of the tool (i.e., ignimed UG) significantly predicted the trustworthiness of the tool itself ($\beta = 0.632, t_{(35)} = 3.573, p = 0.001$): the higher the provider's trustworthiness, the higher the trustworthiness of the tool. Similar, but not significantly, the trustworthiness of the automatic speech recognition predicted the trustworthiness of the tool ($\beta = 0.371, t_{(35)} = 1.768, p = 0.086$): the higher the trustworthiness of the speech recognition, the higher the perceived trustworthiness of the tool. In contrast, the higher the platform provider's (i.e., Amazon) scores, the lower the trustworthiness of the tool, only by trend, however ($\beta = -0.446, t_{(35)} = -1.772, p = 0.085$).

DISCUSSION

Aim of the Present Study

Voice-based (artificial intelligence) systems serving as digital assistants have evolved dramatically within the last few years.

The healthcare sector has been referred to as an emerging market for these systems, which imposes different requirements on the systems than private usage scenarios do. Data revealed here are more personal and more sensitive, resulting in increasing engineering requirements regarding data security, for example. To establish voice-based systems in a more sensitive context, the users' perspective needs to be considered. In a healthcare context, users need to trust their technological counterpart to disclose personal information. However, the trustworthiness of most of the systems in the market and the users' willingness to trust in the applications has not been analyzed yet. The present study bridged this research gap. In an empirical study, the trustworthiness of a voice-based anamnesis tool was analyzed. In two different conditions, participants either interacted with a VA, which was introduced as a "specialist" or a "generalist". Then, they rated the trustworthiness of the tool, distinguishing between different source layers of trust (provider, platform provider, automatic speech recognition in general, data receiver). To ensure external reliability, participants interacted with an anamnesis tool for sexual health, which collected health data by asking questions regarding their puberty, sexual orientation, diseases/hygiene and sexual activity. They were informed that the tool uses artificial intelligence to provide a diagnosis, which would be sent to the gynecologists/urologist.

Answering the Research Questions

The present study investigated three research questions: 1) Does the expert framing of a voice assistant increase its trustworthiness in the context of Further, 2) Does individual dispositions influence the perceived trustworthiness? Finally 3) Do different trustworthiness source indicators (e.g., the assistant tool, the providers, the receiver), interact with each other?

In line with previous studies, the present results revealed that participants, who perceive the VA tool as a specialist tool, reported higher levels of trustworthiness across all different source layers—compared to participants, who perceived the tool as a generalist. Considering, that the tool acted completely identically in both conditions and that the conditions only differed in terms of the introduction of the VA to the participant (written introduction and introduction presented by the VA itself), the present study highlights the manipulability of the users' perception of the system and the effects this perception has on the evaluation of the trustworthiness of the system. The way a diagnostic tool is introduced to the patient seems to be of considerable importance when it comes to the patient's perception of the tool and the willingness to interact with it. As the present study reveals, a few words can fundamentally change the patients' opinions of the tool, which might affect their willingness to cooperate. In the presence of the ongoing worldwide pandemic, we all learnt about the need for intelligent tools, which support physicians with remote anamnesis and diagnosis to unburden the stationary medical offices. Our study shows how important it is to not only consider engineering aspects and ensure that the system functions properly but to consider also the users' perception of the tool and the resulting trustworthiness. Thus, our results offer promising first insights for developers and designers. However, our results also refer to risks. Many health-related tools conquer the

market without any quality checks. If these tools framed themselves as experts or specialists, users could be easily misled.

Following the basic assumption of the media equation approach and its significant research body confirming the idea that social rules and dynamics, which guide human-human interaction, similarly apply to human-computer interaction (Reeves and Nass, 1996), voice-based systems could be regarded as a new era of technological counterpart. Being able to recognize process and produce human language, VA adopt features that have been exclusively human until recently. Consequently, VA can verbally introduce themselves to the users resulting in a powerful manipulation of the users' perception. The presented results show how easily and effectively the impression of a VA can be manipulated. Furthermore, our results indicate an area of research, today's HCI research tends to miss too often. While its primary focus is on the effect of gestalt design on usability and user experience, our results encourage to refer to the users' perspective on the system and the perceived trustworthiness as an essential aspect of a responsible and serious design, which bears chances and risks for both high-quality and low-quality applications.

Referring to methodological challenges, the present study reveals limitations of the way we manipulated the impression of the VA. Participants read an introduction text, which referred to the VA as either a specialist or a generalist tool. Moreover, the VA introduced itself as a specialist or a generalist. Although a pre-test was conducted to ensure the manipulation, not all participants took the hints resulting in participants of the "specialist condition", who did not refer to the VA as a specialist. Similarly, not all participants took the hints resulting in participants of the "generalist condition", who did not refer to the VA as a generalist. Future studies in this area need to conduct a manipulation check to ensure their manipulation or to adopt their analysis strategy (e.g., post-hoc assignments of groups). In our study, unadjusted analyses, which strictly followed the intended manipulation, resulted in reduced effects compared to the newly composed groups following the participants' actual perception of the tool. Additionally, future research should focus on manipulations that are more effective. Following the *source-credibility* model and the *source-attractiveness* model, the perspective on perceived competence is more complex. Besides the perception of expertise, authoritativeness, competence, qualification, or a system perceived as being trained, informed, and educated could contribute to the attribution of competence (Ohanian, 1990; McKnight et al., 2002). Future studies should use the variety of possibilities to manipulate the perceived competence of a VA.

From a theoretical perspective, competence is only one dimension describing the human trustee's main characteristics. Benevolence and integrity of the trustee are also relevant indicators (e.g., Ohanian, 1990; McKnight et al., 2002). In terms of an artificial trustees, performance (analogous to human competence), clarity (analogous to human benevolence), and transparency (analogous to integrity) are further dimension, which determine the impression (Backhaus, 2017). Future studies should widen the perspective and refer to the multiple dimensions. Additionally, human information

processing was introduced to follow two different routes: the systematic (central) or the heuristic (peripheral) route (Cacioppo and Petty, 1986) with personal relevant topics increasing the probability to be processed systematically. Sexual health, the topic of the anamnesis tool of the present study, was shown to be of personal relevance (Kraft Foods, 2009), indicating systematic processing of judgment-relevant cues (e.g., source's expertise) (Cacioppo and Petty, 1986). This might explain the limited effect of the manipulation on the ratings of trustworthiness. Possibly, the personally relevant topic of sexual health triggers the central route of processing resulting in the rather quick labeling of the VA as a heuristic (peripheral) cue to have a limited effect. Moreover, the interaction with the tool might have further diminished the effect of the manipulation. That might also explain why only participants, who explicitly evaluated the tool as a specialist, showed more trustworthiness. Thus, it should be further investigated whether effects of the heuristical design of health-related websites, for example (Gore and Madhavan, 1993), can be transferred to voice-based anamnesis tools assessing highly personal relevant topics.

Regarding the second research question, the present results show only minor effects of the participants' individual characteristics. From the multiple source layers of trustworthiness, the participants' general disposition to trust only impacted the perceived trustworthiness of automatic speech recognition. Possibly, our sample was too homogenous regarding the participants' disposition to trust: mean values of trust-disposition were relatively high ($M = 3.567$ on a 5-point scale) and rather low ($SD = 1.03$). Future studies could consider to incorporate predictors, which are more closely connected with the selected use case such as the tendency to disclose sexual health (Mayer et al., 1995).

The third research question explored the relationship between the different source layers of trustworthiness. When predicting the trustworthiness of the tool, only the tool provider's perceived trustworthiness was a significant predictor. The trustworthiness of the platform provider and automatic speech recognition are related by trend while the data receiver's (gynecologists/urologist) trustworthiness was of minor importance. However, as our participants knew that their data would be only saved on the university's server, the latter results might have been different if the data were transferred to the attending physician (or if participants assumed data transfer). Nevertheless, results are interpreted as a careful first confirmation of Koh and Sundar (2010), who postulate that the perception of an artificial counterpart is not only influenced by the characteristics of the tool itself but also by indicators related to the tool. Referring to today's most popular VAs for private use, Amazon Echo and Google Home, both tools might be closely associated with the perceived image or trustworthiness of the companies. If such consumer products are used in the context of healthcare, reservations regarding the companies might have an impact. Furthermore, the general view of automatic speech recognition affected the perceived trustworthiness of the tool. Thus, current public debates about digitalization or artificial intelligence should also be

considered when designing and using VA for health-related applications.

Limitations and Future Work

To summarize the limitations and suggestions for future work presented above, the manipulation of competence and the additional indicators of trustworthiness need to be reconsidered. Future work might consider more fine-grained and more in-depth operationalization of different expert levels (e.g., referring to the performance of the tool), include further manipulations of the competence dimensions (e.g., referring to a trained system), or incorporate the dimensions of clarity (analogous to human benevolence) or transparency (resample integrity). The perceived trustworthiness might result from a systematic (central) information process due to potential high personal relevance. Future work should investigate whether the effects of the heuristic design of health-related websites, for example (Gore and Madhavan, 1993; Kareklas et al., 2015), can be transferred to voice-based anamnesis tools assessing highly personal relevant topics. The data receiver (gynecologists/urologist) played only a subordinate role in the present study. As participants knew that their disclosed data would be stored on university servers, they were of minor importance. Future work should increase the external validity of the experiment by incorporating the data receiver more explicitly. Finally, only the tool's expert status and not of the additional source layers of trustworthiness have been manipulated, resulting in relatively simple analyses of interaction effects between the trustworthiness indicators. Future studies might choose a more elaborate design. Finally, perceived trustworthiness is an essential topic for different application areas such as education (Troussas et al., 2021). Another important field might be the perceived trustworthiness of multilingual voice assistants applied in multilingual societies (Mukherjee et al., 2021). A different approach to the perceived trustworthiness would be testing the impact of different dialogue architectures (e.g., Fernández et al., 2005). Strategies of dialogue design can be very different and impact on user's trustworthiness. Future studies should investigate if hardcoded intents or flexible and natural spoken interactions have a different impact.

Conclusion and Contribution

Voice assistants gain in importance in healthcare contexts. With remote anamnesis and diagnoses gaining in importance these days, voice-based systems offer promising contributions, for instance, in the area of medical diagnoses. Using voice assistants in data sensitive contexts draws attention to the concept of trust: if patients were to reveal personal, sensitive information to the voice-based systems, they would need to trust them. However, the analysis and the understanding of the psychological processes characterizing the patient-voice assistant interaction is still in their early stages. For human-human relationships, psychological research revealed the characteristics of individuals, who give trust (trustor) and those who receive trust (trustee). Moreover, research established models of the characteristics, which are processed and attributed (e.g.,

source-credibility model, source-attractiveness model, HSM, ELM). Researchers in the field of human-computer interaction transferred this knowledge to interactions with technological counterparts (e.g., television, web pages, web agents). However, little is known about voice-based tools, which have become increasingly popular, and which involve more complex, more humanlike features (speech processing) compared to technology so far. The present study contributes to close this research gap by presenting ideas for the design of VAs, which have been derived from literature. Furthermore, the study provides empirical data of human users interacting with a device to disclose health-related information. Results showed that participants, who perceived the VA tool as a specialist tool, reported higher trustworthiness scores than participants, who thought to interact with a generalist tool. To conclude, the users' perception significantly influences the trust users have in the VA. Furthermore, influencing this perception was shown to be rather easy: a short-written introduction and a "spoken" introduction presented by the VA itself were sufficient to affect the users' perception and their trust in the system significantly. In sum, we want to draw attention to the importance of the human user's perspective when interacting with technology. Future studies need to address the trustworthiness of technology to contribute to more responsible and serious design processes to take the chances technology offers and to avoid the risks of low-quality applications.

REFERENCES

- Backhaus, N. (2017). User's Trust and User Experience in Technical Systems: Studies on Websites and Cloud Computing 1–281. doi:10.14279/depositonce-5706
- Bär, N. (2014). Human-Computer Interaction and Online Users' Trust. PhD dissertation. TU Chemnitz. Retrieved from: <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa149685>
- Bär, N., Hoffmann, A., and Krems, J. (2011). "Entwicklung von Testmaterial zur experimentellen Untersuchung des Einflusses von Usability auf Online-Trust," in *Reflexionen und Visionen der Mensch-Maschine-Interaktion – Aus der Vergangenheit lernen, Zukunft gestalten*. Editors S. Schmid, M. Elepfandt, J. Adenauer, and A. Lichtenstein, 627–631.
- Bates, B. R., Romina, S., Ahmed, R., and Hopson, D. (2006). The Effect of Source Credibility on Consumers' Perceptions of the Quality of Health Information on the Internet. *Med. Inform. Internet Med.* 31, 45–52. doi:10.1080/14639230600552601
- Baumeister, J., Sehne, V., and Wiennrich, C. (2019). "A Systematic View on Speech Assistants for Service Technicians," in *LWDA Jäschke Robert and Weidlich Matthias Editors*. Berlin, Germany, 195–206. doi:10.1136/bmjspcare-2019-huknc.228
- Beierlein, C., Kemper, C. J., Kovaleva, A., and Rammstedt, B. (2012). "Kurzsкала zur Messung des zwischenmenschlichen Vertrauens: Die Kurzsкала Interpersonales Vertrauen (KUSIV3)[Short scale for assessing interpersonal trust: The short scale interpersonal trust (KUSIV3)]," in *GESIS Working Papers 2012|22*, Köln.
- Cacioppo, J. T., and Petty, R. E. (1986). in "He Elaboration Likelihood Model of Persuasion," in *Advances in Experimental Social Psychology*. Editor L. Berkowitz (New York: Academic Press), 123–205.
- Casaló, L. V., Flavián, C., and Guinalíu, M. (2007). The Role of Security, Privacy, Usability and Reputation in the Development of Online Banking. *Online Inf. Rev.* 31 (5), 583–603. doi:10.1108/14684520710832315
- Chaiken, S., and Maheswaran, D. (1994). Heuristic Processing Can Bias Systematic Processing: Effects of Source Credibility, Argument Ambiguity, and Task

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CW: concept of study, theoretical background, argumentation line, data analyses, supervision; CR: skill implementation, study conductance, data analyses; AC: theoretical background, argumentation line, supervision.

FUNDING

This publication was supported by the Open Access Publication Fund of the University of Wuerzburg.

Importance on Attitude Judgment. *J. Personal. Soc. Psychol.* 66, 460–473. doi:10.1037/0022-3514.66.3.460

Chaiken, S. (1987). "The Heuristic Model of Persuasion," in *Social influence: the ontario symposium* M. P. Zanna, J. M. Olson, and C. P. Herman Editors. Lawrence Erlbaum Associates, Inc, 5, 3–39.

Clark, L. V. (1987). *Sexual Self-Disclosure to Parents and Friends*. Texas: University Press.

Corritore, C. L., Kracher, B., and Wiedenbeck, S. (2003). On-line Trust: Concepts, Evolving Themes, a Model. *Int. J. Human-Computer Stud.* 58, 737–758. doi:10.1016/s1071-5819(03)00041-7

Crisci, R., and Kassino, H. (1973). Effect of Perceived Expertise, Strength of Advice, and Environmental Setting on Parental Compliance. *J. Soc. Psychol.* 89, 245–250. doi:10.1080/00224545.1973.9922597

Eisend, M. (2006). Source Credibility Dimensions in Marketing Communication—A Generalized Solution. *J. Empir. Gen. Mark. Sci.* 10.

Endreß, M. (2010). "Vertrauen—soziologische Perspektiven," in *Vertrauen—zwischen Sozialem Kitt Senkung Von Transaktionskosten*, 91–113.

ePharmaINSIDER (2018). Das sind die Top 12 Gesundheits-Chatbots. (Accessed March 17, 2021). Available at: <https://www.epharmainsider.com/die-top-12-gesundheits-chatbots/>.

Fernández, F., Ferreiros, J., Sama, V., Montero, J. M., Segundo, R. S., Macías-Guarasa, J., et al. (2005). "Speech Interface for Controlling an Hi-Fi Audio System Based on a Bayesian Belief Networks Approach for Dialog Modeling," in *Ninth European Conference on Speech Communication and Technology*.

Giffin, K. (1967). The Contribution of Studies of Source Credibility to a Theory of Interpersonal Trust in the Communication Process. *Psychol. Bull.* 68, 104–120. doi:10.1037/h0024833

Gore, P., and Madhavan, S. S. (1993). Credibility of the Sources of Information for Non-prescription Medicines. *J. Soc. Adm. Pharm.* 10, 109–122.

Hernandez, A. (2021). The Best 7 Free and Open Source Speech Recognition Software Solutions. Goodfirms. Available at: <https://www.goodfirms.co/blog/best-free-open-source-speech-recognition-software> (Accessed March 17, 2021).

- Hoff, K. A., and Bashir, M. (2015). Trust in Automation. *Hum. Factors* 57, 407–434. doi:10.1177/0018720814547570
- Hörner, T. (2019). "Sprachassistenten im Marketing," in *Marketing mit Sprachassistenten*. Springer, Gabler: Wiesbaden, 49–113. doi:10.1007/978-3-658-25650-0_3
- Hovland, C. L., Janis, I. L., and Kelley, H. H. (1953). *Communication and Persuasion*. New Haven, CT: Yale University Press
- idealo (2020). E-Commerce-Trends 2020: Millennials Treiben Innovationen Voran. Idealo. Available at: <https://www.idealo.de/unternehmen/pressemitteilungen/ecommerce-trends-2020/> (Accessed March 17, 2021).
- Kareklas, I., Muehling, D. D., and Weber, T. J. (2015). Reexamining Health Messages in the Digital Age: A Fresh Look at Source Credibility Effects. *J. Advertising* 44, 88–104. doi:10.1080/00913367.2015.1018461
- Kim, K. J. (2014). Can Smartphones Be Specialists? Effects of Specialization in mobile Advertising. *Telematics Inform.* 31, 640–647. doi:10.1016/j.tele.2013.12.003
- Kim, K. J. (2016). Interacting Socially with the Internet of Things (IoT): Effects of Source Attribution and Specialization in Human-IoT Interaction. *J. Comput-mediated Comm.* 21, 420–435. doi:10.1111/jcc4.12177
- Koh, Y. J., and Sundar, S. S. (2010). Heuristic versus Systematic Processing of Specialist versus Generalist Sources in Online media. *Hum. Commun. Res.* 36, 103–124. doi:10.1111/j.1468-2958.2010.01370.x
- Kraft Foods (2009). Tabuthemen. Statista. Available at: <https://de.statista.com/statistik/daten/studie/4464/umfrage/themen-ueber-die-kaum-gesprochen-wird/> (Accessed March 17, 2021). doi:10.1007/978-3-8349-8235-3
- Leshner, G., Reeves, B., and Nass, C. (1998). Switching Channels: The Effects of Television Channels on the Mental Representations of Television News. *J. Broadcasting Electron. Media* 42, 21–33. doi:10.1080/08838159809364432
- Liew, T. W., and Tan, S.-M. (2018). Exploring the Effects of Specialist versus Generalist Embodied Virtual Agents in a Multi-Product Category Online Store. *Telematics Inform.* 35, 122–135. doi:10.1016/j.tele.2017.10.005
- Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Amr* 20, 709–734. doi:10.5465/amr.1995.9508080335
- McKnight, D. H., Choudhury, V., and Kacmar, C. (2002). Developing and Validating Trust Measures for E-Commerce: An Integrative Typology. *Inf. Syst. Res.* 13, 334–359. doi:10.1287/isre.13.3.334.81
- McKnight, D. H., Cummings, L. L., and Chervany, N. L. (1998). Initial Trust Formation in New Organizational Relationships. *Amr* 23, 473–490. doi:10.5465/amr.1998.926622
- Meticulous Market Research (2021). Healthcare Virtual Assistant Market By Product (Chatbot And Smart Speaker), Technology (Speech Recognition, Text-To-Speech, And Text Based), End User (Providers, Payers, And Other End User), And Geography - Global Forecast To 2025. Available at: <http://www.meticulousresearch.com/> (Accessed March 17, 2021).
- Mukherjee, S., Nediyanath, A., Singh, A., Prasan, V., Gogoi, D. V., and Parmar, S. P. S. (2021). "Intent Classification from Code Mixed Input for Virtual Assistants," in 2021 IEEE 15th International Conference on Semantic Computing (ICSC) (IEEE), 108–111.
- Nass, C., and Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *J. Soc. Issues* 56, 81–103. doi:10.1111/0022-4537.00153
- Ohanian, R. (1990). Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness. *J. Advertising* 19, 39–52. doi:10.1080/00913367.1990.10673191
- Pornpitakpan, C. (2004). The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *J. Appl. Soc. Psychol* 34, 243–281. doi:10.1111/j.1559-1816.2004.tb02547.x
- Reeves, B., and Nass, C. (1996). *The media Equation: How People Treat Computers, Television, and New media like Real People and Places*. Cambridge, UK: Cambridge University Press.
- Riedl, B., Gallenkamp, J., and Hawaii, A. P. (2013). "The Moderating Role of Virtuality on Trust in Leaders and the Consequences on Performance," in 2013 46th Hawaii International Conference on System Sciences. IEEE, 373–385. Available at: <https://ieeexplore.ieee.org/Xplore/home.jsp>. doi:10.1109/hicss.2013.644
- Söllner, M., Hoffmann, A., Hoffmann, H., and Leimeister, J. M. (2012). Vertrauensunterstützung für sozio-technische ubiquitäre Systeme. *Z. Betriebswirtsch* 82, 109–140. doi:10.1007/s11573-012-0584-x
- The Medical Futurist (2020). The Top 12 Healthcare Chatbots. *Med. Futur.* Available at: <https://medicalfuturist.com/top-12-health-chatbots> (Accessed March 17, 2021).
- Troussas, C., Krouska, A., Alepis, E., and Virvou, M. (2021). Intelligent and Adaptive Tutoring through a Social Network for Higher Education. *New Rev. Hypermedia Multimedia*, 1–30. doi:10.1080/13614568.2021.1908436
- Venkatraman, V., Dimoka, A., Pavlou, P. A., Vo, K., Hampton, W., Bollinger, B., et al. (2015). Predicting Advertising success beyond Traditional Measures: New Insights from Neurophysiological Methods and Market Response Modeling. *J. Marketing Res.* 52, 436–452. doi:10.1509/jmr.13.0593
- McKnight, D. H., Choudhury, V., and Kacmar, C. (2002b). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research* 13, 334–359. doi:10.1287/isre.13.3.334.81

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wienrich, Reitelbach and Carolus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.