HUMAN-MACHINE COMMUNICATION

Human-Machine Communication
Volume 4, 2022
https://doi.org/10.30658/hmc.4.5

Human, Hybrid, or Machine? Exploring the Trustworthiness of Voice-Based Assistants

Lisa Weidmüller[1]

1 Science and Technology Communication, Technical University Dresden, Dresden, Germany

Abstract

This study investigates how people assess the trustworthiness of perceptually hybrid communicative technologies such as voice-based assistants (VBAs). VBAs are often perceived as hybrids between human and machine, which challenges previously distinct definitions of human and machine trustworthiness. Thus, this study explores how the two trustworthiness models can be combined in a hybrid trustworthiness model, which model (human, hybrid, or machine) is most applicable to examine VBA trustworthiness, and whether this differs between respondents with different levels of prior experience with VBAs. Results from two surveys revealed that, overall, the human model exhibited the best model fit; however, the hybrid model also showed acceptable model fit as prior experience increased. Findings are discussed considering the ongoing discourse to establish adequate measures for HMC research.

Keywords: voice-based assistant, trustworthiness, trust, hybrid, scale, survey, prior experience

Introduction

In Human-Machine Communication (HMC) research, we are dealing with interactions between humans and perceptually *hybrid* communicative technologies such as voice-based assistants (VBAs). This means that humans who communicate and interact with these

CONTACT Lisa Weidmüller • Science and Technology Communication • Technical University Dresden • Zellescher Weg 17 • 01069 Dresden, Germany • lisa.weidmueller@tu-dresden.de

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

technologies often do not perceive them solely as machines. Instead, because machines are increasingly built and programmed to imitate human capabilities and behavior "by exchanging messages with people or by performing a communicative task on their behalf" (Guzman, 2020, p. 37), they exhibit social cues that may prompt the attribution of human traits or social behavior to them (A. P. Edwards, 2018; Gambino et al., 2020; Garcia et al., 2018; Nass & Moon, 2000; Reeves & Nass, 1996). In fact, humans often perceive these technologies as "social things" (Guzman, 2015) or "personified things" (Etzrodt & Engesser, 2021), ascribing both human and machine characteristics to them.

Because of this perceptual hybridity, we need to adapt the way we investigate interactions with these communicative technologies and their effects (Chita-Tegmark et al., 2021; J. Edwards & Sanoubari, 2019). One context in which this is particularly relevant is the formation of trust toward communicative technologies. Trust has been identified as a crucial factor in the use and adoption of technology (e.g., Gao & Bai, 2014; C.-J. Lee et al., 2005; Priest, 2001). However, which trustworthiness characteristics form the basis for trust development depends on the perceived nature of the trustee, which is understood to differ substantially between human and machine trustees (Akter et al., 2011; Colquitt et al., 2007; Lankton et al., 2015; J. D. Lee & See, 2004). Therefore, communicative technologies like VBAs that blur perceptual categories, such as human and machine, are especially challenging.

VBAs, like Amazon's Alexa, Apple's Siri, or Google Assistant, represent a communicative technology that exhibits primary social cues due to its human-like conversational user interface (CUI) (Lombard & Xu, 2021). This triggers social reactions, interpersonal interaction patterns, and the attribution of (social) agency (Burgoon et al., 1999; Gambino et al., 2020; Lombard & Xu, 2021; McTear et al., 2016). Increasingly, VBAs also adapt a communicative role similar to that of a traditional news anchor. By selecting news content from external sources (e.g., media companies) and presenting the stories in their own voice and character in answer to a search query, VBAs take on the role of *artificial news anchors*. Their trustworthiness is therefore of particular importance because people may rely on the information they receive from VBAs to form opinions about the world surrounding them. As more people use VBAs to learn about the news (Ammari et al., 2019; Kinsella & Mutchler, 2020; Natale & Cooke, 2020; Newman, 2018) and VBA developer companies work to expand their respective VBAs' abilities to present news (Lyons, 2020; Porter, 2019), this use context, including VBAs' related trustworthiness, may deserve more attention than it currently receives (J. Edwards & Sanoubari, 2019). However, in this context, there currently exists no trustworthiness measure that adequately recognizes VBAs' perceptually hybrid nature as trustees and communicators.

Thus, the purpose of this study is to (1) empirically explore how the distinct human and machine trustworthiness models can be combined in a hybrid trustworthiness model for VBAs as artificial news anchors, (2) examine which model (human, hybrid, or machine) is most applicable to examine VBA trustworthiness in this context, and (3) determine whether this differs between respondents with different levels of prior experience with the VBA (i.e., no prior experience, indirect experience, or direct experience).

For these purposes, the paper, first, explicates what makes VBAs perceptually hybrid trustees and discusses what this means for assessing VBAs' trustworthiness. Second, a hybrid model for VBA trustworthiness is exploratively developed by using data from two online surveys. Third, the competing human, machine, and hybrid models are empirically tested and compared to answer the following questions:

> **(RQ1)** Which trustworthiness model (human, hybrid, or machine) exhibits the best model fit and is thus most applicable to investigate VBA trustworthiness?
>
> **(RQ2)** How does people's level of prior experience with the VBA affect which trustworthiness model is most applicable?

Finally, implications of the findings for HMC research are discussed, and limitations, as well as directions for future research, are presented.

The Perceptual Hybridity of Voice-Based Assistants

VBAs can be distinguished from previous assistance applications by their sophisticated voice interface and dialogue system (Yang et al., 2019). Based on automatic speech recognition, analysis (natural language processing), synthesis (text-to-speech), and artificial intelligence (AI), VBAs can recognize and understand spoken instructions after receiving a wake word and can give meaningful answers or present information relevant to a query in a human-sounding voice (Deloitte, 2018; McTear et al., 2016). They are thereby able to mirror human interaction patterns, exhibit primary social cues, and simulate intentions through effective and meaningful behavior (Burgoon et al., 1999; Hearst, 2011; Lombard & Xu, 2021; Nass & Moon, 2000). This illusion of a human communicator is assisted by scripted small talk responses that are designed to convey a unique character or persona for each VBA (Natale, 2021), challenging the perception of VBAs as mere machines.

Studies have shown that the VBAs' human-like CUI activates scripts of interpersonal communication (Burgoon et al., 1999; McTear et al., 2016; Moon et al., 2016), prompting users to react socially and attribute human-like traits, like gender or personality, to the VBAs (Etzrodt & Engesser, 2021; Garcia et al., 2018; Guzman, 2019). Scholars have explained this behavior toward machines as a result of people's doubt (e.g., Reeves & Nass, 1996) about the machines' ontological classification. In essence, it is difficult for people to determine "who" or "what" a VBA is (Gunkel, 2020), or whether a VBA is a thing/object or person/subject (Etzrodt & Engesser, 2021). Most notably, people are in doubt about whether a VBA is a *human* or *machine* (Guzman, 2020). Recently, Etzrodt and Engesser (2021) explored the nature of this doubt, uncovering that, rather than assimilating VBAs into one of the schemes, people accommodated their schemes by classifying VBAs as "personified things." In other words, people ascribed both human and machine characteristics to VBAs, though not in the same amount (Etzrodt & Engesser, 2021). Similar results have been found for personified robots, which were attributed—to a certain degree—mental states, sociality, and even morality (Kahn, Jr. et al., 2011).

According to the Oxford English Dictionary (Oxford University Press, n.d.), something that is a mixture of (at least) two different elements is *hybrid*. Thus, VBAs can be understood as perceptual hybrids on the borderline of the human-machine distinction. They thereby challenge existing models of trustworthiness and trust, which are suggested to differ substantially when dealing with a machine(-like) versus a human(-like) trustee (Lankton et al., 2015; J. D. Lee & See, 2004).

Trustworthiness and Trust Development

Trust is relevant in any situation where two or more distinct parties have to rely on one another to successfully complete a task or interaction involving uncertainty, unequally distributed knowledge, and/or the risk of negative consequences (Akter et al., 2011; J. D. Lee & See, 2004; Mayer et al., 1995; McKnight et al., 2011). Trust is therefore not limited to interpersonal exchanges but also includes human-machine interactions, and human-organization interactions (Schaefer et al., 2016). While the basic definition for trust is similar across these interactions, there are also differences, especially between human-human trust, and human-machine trust.

Trust can be defined, both in human-human interactions and human-machine (e.g., robot, automation, or agent) interactions, as an *attitude* of someone (hereafter, trustor) toward something or someone else (hereafter, trustee), coupled with the expectation that relying on the other party will prompt favorable outcomes (Blöbaum, 2016; Colquitt et al., 2007; J. D. Lee & See, 2004; Rousseau et al., 1998). Therefore, trust is not an action or behavior itself, but it can lead to behavioral intentions from which trust-related behavior (e.g., interaction, cooperation, or reliance) may result.

Over time, trust develops based on *trusting beliefs*, which are the result of the trustor's *experience from prior interactions* with a trustee and *immediately perceivable information* about the trustee's nature (J. D. Lee & See, 2004; Mayer et al., 1995; Rousseau et al., 1998). These trusting beliefs are often studied under the term *trustworthiness.* Trustworthiness can be defined as the trustor's attribution and evaluation of a trustee's abilities and characteristics that, the trustor believes, will lead to a beneficial outcome in a *specific context* (J. D. Lee & See, 2004; Mayer et al., 1995; McKnight et al., 2002).

The Role of the Trustee's Nature

The primary difference between human-human trust, and human-machine trust is the *nature of the trustees*. Unlike in human-human interactions, where both interaction partners belong to the same ontological category, in human-machine interactions, they do not (Guzman, 2020), making it more difficult for human trustors to determine the nature of their machine interaction partner. Because the nature of a trustee determines which of the trustee's characteristics are relevant to assess trustworthiness (Lankton et al., 2015; J. D. Lee & See, 2004), the perceptual hybridity of VBAs between human and machine challenges previously distinct definitions of human and machine trustworthiness.

A widely acknowledged definition of human-human trustworthiness—which guides the understanding of human-like trustworthiness in this study—is based on the following three dimensions: *integrity*, *competence,* and *benevolence* (Blöbaum, 2016; Mayer et al., 1995). To assess a human(-like) trustee as trustworthy thereby means that the trustee is perceived to adhere to certain moral and ethical values that are important to the trustor, have the skills and characteristics to fulfill a certain task skillfully, and intent to do good to the trustor.

On the other hand, a widely acknowledged definition of trust in technology, or system-like trust—which guides the understanding of machine-like trustworthiness in this study—is based on the following three dimensions: *reliability*, *functionality,* and *helpfulness* (Chita-Tegmark et al., 2021; Lankton et al., 2015; McKnight et al., 109

). To assess a machine(-like) trustee as trustworthy thereby means that the trustee is perceived to consistently operate properly, have the functions or features needed to fulfill a specific task, and provide adequate help and assistance to the trustor.

These distinct definitions are based on the assumption that human(-like) trustees have the power to choose and make ethical decisions (i.e., intentions and volition), while machine(-like) trustees do not—apart from how they are preprogrammed (Lankton et al., 2015; J. D. Lee & See, 2004; Mayer et al., 1995). Thus, human trustees can choose to adhere to certain values or do good to the trustor, or they can choose not to. Machine trustees, on the other hand, cannot choose to operate properly or provide adequate assistance. Instead, machines need to have been correctly designed by human programmers to function without error and to provide the help needed by the user—without any intentionality on the part of the machine being involved. However, research shows that users can attribute intentionality to a technology itself (Guzman, 2019; Reeves & Nass, 1996; Sundar & Nass, 2000), especially when the technology exhibits sophisticated social cues that imply human-likeness (Gambino et al., 2020; Nass & Moon, 2000). Therefore, human and machine trustworthiness dimensions have to be combined to investigate the trustworthiness of such perceptually hybrid trustees (Chita-Tegmark et al., 2021; J. Edwards & Sanoubari, 2019).

The Role of Context

In addition to the differences between human and machine trustworthiness, which are based on the *nature of the trustee*, trustworthiness is *domain specific* (Mayer et al., 1995). That means that the goal or context of the interaction between the trustor and trustee determines which of the trustee's characteristics are relevant for the trustor's assessment of the trustee's trustworthiness (J. D. Lee & See, 2004).

In the context of presenting news, trustworthiness has been widely studied under the term *credibility* regarding the three domains of source, medium, and message. For investigating VBA trustworthiness, both the definitions of source and media credibility are relevant because the VBAs' human-like CUI has also been found to cause uncertainty about their role in the communication process. Some users, for instance, are inclined to identify VBAs as veritable sources rather than mediating channels of a message (Guzman, 2019). This is in line with the CASA (*Computers Are Social Actors*) paradigm, which purports that technologies that exhibit sufficient social cues—such as a human-sounding voice—can be perceived as the *source* of communication (Gambino et al., 2020; Lombard & Xu, 2021; Nass & Moon, 2000; Sundar & Nass, 2000).

Both the definitions of source and media credibility consistently include the dimension of expertise, defined as the *ability to know the truth*, and a dimension called trustworthiness, defined as the *motivation to tell the truth* (e.g., Hovland et al., 1953; Metzger et al., 2003). Thus, both credibility dimensions emphasize the importance of *truthfulness*, which relates to the credibility assessment of the message. Since this study investigates how people assess the trustworthiness of VBAs in the context of presenting news, the human-like and machine-like trustworthiness dimensions need to be adapted to this specific context by incorporating the characteristic of truthfulness and the reference to the presented message.

VBAs as Hybrid Trustees

As aforementioned, research has shown that VBAs are perceptual hybrids. While undeniably technological devices, their human-like CUI causes uncertainty about whether VBAs are ontologically human or machine and, consequently, people attribute both human and machine characteristics to them (e.g., Etzrodt & Engesser, 2021). Additionally, this causes an overlap between the attributed role in the communication process as source or channel (e.g., Guzman, 2019). To incorporate this perceptual hybridity when examining VBA trustworthiness, this study explores how a hybrid trustworthiness model as a mix of human(-like) and machine(-like) trustworthiness attributes adapted to the news presentation context might look. Therefore, an exploratory factor analysis (EFA) will be conducted with items representing all six trustworthiness dimensions (human: integrity, competence, benevolence; machine: reliability, functionality, helpfulness). Theoretically, it would be possible that in a hybrid trustworthiness model, all six dimensions must be retained, or that two dimensions—a human and a machine trustworthiness dimension with the respective items—could evolve. However, it would also be possible that the characteristics from different dimensions of human and machine trustworthiness could mix and form hybrid dimensions of trustworthiness. The number of such hybrid dimensions is also open, although a three-dimensional structure would correspond to the established three-dimensional structure for both human and machine trustworthiness.

After a hybrid model—of whatever form—is developed, the model fit of the competing models will be compared to investigate *which model is most applicable to investigate VBA trustworthiness (RQ1).*

The Moderating Role of Prior Experience

The literature suggests that prior experience plays an important role in establishing trust toward *any* trustee (J. D. Lee & See, 2004; Mayer et al., 1995; McKnight et al., 2002; Rousseau et al., 1998; Schaefer et al., 2016). Thus, we can distinguish between *initial trust* in an unfamiliar trustee (e.g., when the trustor first interacts with a trustee) and trust that is established based on prior interactions and experience with a trustee (Li et al., 2008; McKnight et al., 2002). Without having firsthand experience with a trustee, a trustor can still form an initial trustworthiness impression based on past experiences with similar trustees (Schaefer et al., 2016) or indirect experiences with the trustee through secondhand information from organizational and cultural contexts (J. D. Lee & See, 2004). Therefore, indirect experience with VBAs through advertisements, fiction, news, or gossip from others may also inform people's expectations of VBAs and thus their trustworthiness assessment.

Previous research uncovered that people's expectations (i.e., mental models) of technologies that exhibit social cues change with growing experience (A. P. Edwards et al., 2019;

Gambino et al., 2020; Horstmann & Krämer, 2019; Lankton et al., 2014). Consequently, it is likely that people who have no prior experience with a VBA (i.e., those who can neither draw upon firsthand nor secondhand experience), those with indirect (secondhand) prior experience, and those with direct (firsthand) prior experience have different expectations of VBAs and thus assess their trustworthiness differently. For example, those with no prior experience or indirect experience possibly overestimate the human-likeness of VBAs and, thus, the human model could be most applicable. For the first group, this could simply be due to a novelty effect, for the latter it could be due to secondhand information from fiction or advertisements, where the human-likeness of such technologies is often emphasized. People with direct experience on the other hand, may have already adapted their mental models and, thus, the hybrid trustworthiness model might be most applicable. This will be examined with the following research question:

> **(RQ2)** How does people's level of prior experience with the VBA affect which trustworthiness model is most applicable?

Method

Sample

To investigate these research questions, an online survey was conducted in late 2018. Respondents were recruited among the students of a large German university via the university's email list. After eliminating incomplete cases (dropout rate: 2.7%) and cases of low quality according to the relative speed of completion and percentage of missing answers (Leiner, 2019), the final sample consisted of N = 853 students (response rate: 2.6%). On average, the respondents were 23 years old (SD = 4.81, age range = 17–50), and they were almost evenly split between men (52%) and women (48%). Most of the respondents were undergraduates (76%), and the vast majority already knew the VBA to which they were randomly assigned in the survey (Alexa: 96%, Google Assistant: 72%). However, for both VBAs, the students' knowledge primarily stemmed from indirect sources such as advertisements, other people, the media, or fiction. Only a few said they owned the assigned VBA themselves (Alexa: 7%, Google Assistant: 33%).

To validate the findings from this pilot study, the survey was additionally conducted with a slightly older sample. Therefore, staff members of the same university were recruited via the university's staff email list (dropout rate: 7.4%, response rate: 6.2%). On average, staff respondents (N = 435, 53% male) were 10 years older (mean age = 33, SD = 10.35, age range = 18–65) than respondents in the student sample. Furthermore, 69% of staff respondents were graduates compared to only 24% in the student sample. Both samples therefore represent two slightly different educational phases: those still in higher education (students) versus those who have finished higher education (staff). Regarding prior experience with the VBAs, the staff sample largely resembled the student sample, but fewer knew their assigned VBA (Alexa: 93%, Google Assistant: 65%) or owned the VBA (Alexa: 6%, Google Assistant: 24%). Both samples will be used to validate the exploratively derived hybrid trustworthiness model and account for possible cohort effects.

Procedure

Because the level of VBA adoption in Germany was still low at the time of the survey, with only 26% of German internet users using VBAs in general (Taş et al., 2019) and only 5% using smart speakers (Newman et al., 2018), a *demonstrational survey design* was chosen. After asking about respondents' prior experience with several VBAs, they were shown prerecorded videos of either the smart speaker variant of Google Assistant (Google Home) or Alexa (Amazon Echo) within the online survey. These two VBAs were selected due to their leading market position (Kinsella & Mutchler, 2020; Statista, 2021); however, the random assignment to either one of the VBAs is not essential for the paper at hand. The videos were between 11 and 19 seconds long and were presented in a way that simulated interactions with the respective VBA in a news use scenario. All respondents successively activated three predefined, news-related questions by clicking a button. Then, they received the VBAs' answers in the form of the prerecorded videos, which are available in this study's OSF repository (including English transcripts).

Though this presentation mode limits the actual conversational nature of interacting with VBAs, it was chosen for three reasons. First, while both VBAs were already available on the German market, making direct or indirect experience with them possible, adoption in Germany was still low. Thus, the prerecorded videos allowed respondents to get a good impression of how VBAs present news.[1] Second, VBA functionality for presenting news was still in its infancy in Germany at the time of the survey. Thus, the prerecorded videos made it possible to simulate a functionality, which was already a reality in English-speaking countries, while still using VBAs that were available on the German market. Third, VBAs present different content depending on which VBA is used, the time and date, and how a search query is formulated. The use of predefined questions and prerecorded videos was therefore necessary to ensure that all respondents received the same content, no matter which VBA they were assigned or when they participated in the survey. The content was selected from real news of quality German media, which the VBAs were manipulated to present verbally, without citing the source, by using the IFTT ("If This Then That")-App and then recording the answers. For the topic, the implementation of the General Data Protection Regulation (GDPR) was chosen due to its societal relevance in Germany at the time of the study. Additionally, the topic was associated with a certain level of risk and uncertainty because of possible (negative) consequences if the regulations were not implemented correctly. As risk contributes to the relevance of trust(worthiness) (e.g., J. D. Lee & See, 2004; McKnight et al., 2002), the topic provided respondents with the necessary incentive to assess trustworthiness, which respondents were asked to do after they saw the videos.

Measures

VBA Trustworthiness

VBA trustworthiness was measured by asking respondents to indicate how much they agreed with 16 items on a seven-point Likert scale (1 = "Do not agree at all" to 7 =

1. Four additional prerecorded videos that contained the VBAs' answers to predefined "personal" questions (e.g., "How are you?") were included in the beginning of the survey (after asking about prior experience) so that respondents could get a good impression of the respective VBA's persona, as well.

"Agree completely"). Additionally, they were able to choose a no response option: "I cannot assess that."

Items were developed to represent the definitions of the three human-like trustworthiness dimensions: integrity, competence, and benevolence (Mayer et al., 1995), and the three machine-like trustworthiness dimensions: functionality, reliability, and helpfulness (Lankton et al., 2015; McKnight et al., 2011). Most of the items are based on existing measures used to investigate trust in technology (Fink, 2014; Lankton et al., 2015), e-commerce (Koh & Sundar, 2010; McKnight et al., 2002; Wang & Benbasat, 2016), anthropomorphic agents (Burgoon et al., 1999), human news anchors, journalists (Newhagen & Nass, 1989), and human endorsers (Ohanian, 1990). However, the wording of the items was specifically adapted to the trustee (VBAs) and the context of information presentation.

Human-Like Trustworthiness Items. For integrity, respondents were asked to assess whether the VBA was *trustworthy* and *credible* in general and whether the VBA *gives truthful information*. To be influential within the news presentation context (competence), the VBA had to be assessed as being *competent* and *qualified to provide information*, as well as *reliable*. Benevolence of the VBA was operationalized as *acting in the trustor's interest*, being *interested in the trustor's well-being*, and *providing help if needed*.

Machine-Like Trustworthiness Items. When developing the machine-like trustworthiness items, previous scales were consulted (Lankton et al., 2015; McKnight et al., 2011; Ullman & Malle, 2018). However, the items had to be adapted to the specific trustee and trusting context. Specifically, to measure whether respondents believed that the VBA consistently operates properly, (reliability) items included were *is reliable*, *gives security in information search,* and *respects people's privacy*, thereby incorporating common user (and designer) concerns regarding VBAs (Clark et al., 2019; Easwara Moorthy & Vu, 2015; Kinsella & Mutchler, 2020; Lei et al., 2018; Olson & Kemery, 2019). Items for the VBAs' functionality to provide information were based on aspects of understandability (*understandable and fluent presentation*) and ease of use (*intelligent* and *well-programmed*). Helpfulness was defined as providing help and advice necessary to fulfill a task, which in this context means selecting and presenting relevant news in response to a query. Therefore, respondents had to indicate whether they believed the VBA *is useful*, *gives relevant answers*, and *provides help if needed*.

An overview of the items' wordings and the underlying definitions is given in Table 1.[2] All of the items will serve as the basis for the exploratory development of the hybrid trustworthiness model.

2. Conceptually, two items belong to both a human-like and a machine-like trustworthiness dimension. First, to be *reliable* was developed as a competence characteristic within the domain of news presentation (human-like: competence), and to reflect the characteristic to consistently operate properly (machine-like: reliability). Second, the definitions of benevolence (human-like) and helpfulness (machine-like) share the similar notion of *providing help if needed*.

TABLE 1 Definition and Operationalization of Human- and Machine-Like Trustworthiness Dimensions Adapted to the Context of News Presentation

Human-Like Trustworthiness			Machine-Like Trustworthiness		
Dimensions		**Items**	**Dimensions**		**Items**
Integrity	The belief that a trustee adheres to a set of principles that the trustor finds acceptable *(domain-specific)*	• can be trusted • is credible • gives truthful information	**Reliability**	The belief that the specific technology will consistently operate properly	• is reliable • provides security in information search • respects people's privacy
Competence	The belief that the trustee has the skills, competencies, and characteristics needed to have influence within a *specific domain*	• is reliable • is competent in providing information • is qualified to provide information	**Functionality**	The belief that the specific technology has the capability, functionality, or features to do for one what one needs to be done *(task-specific)*	• answers understandably & fluently • is well-programmed • is intelligent
Benevolence	The belief that the trustee will want to do good to the trustor apart from an egocentric motive	• acts in my best interest • is interested in my well-being • would do its best to help me if I needed help	**Helpfulness**	The belief that the technology provides adequate, effective, and responsive help or advice necessary to successfully complete a *specific task*	• is useful • gives relevant answers • would do its best to help me if I needed help

Notes. Based on Burgoon et al., 1999; Fink, 2014; Koh & Sundar, 2010; Lankton et al., 2015; Mayer et al., 1995; McKnight et al., 2002; McKnight et al., 2011; Newhagen & Nass, 1989; Ohanian, 1990; Wang & Benbasat, 2016.
Items were translated from German original (see supplemental material).

Prior Experience

Prior experience with the respective VBA was measured by asking whether respondents knew the assigned VBA before the study (yes/no) and from where they had learned about the VBA (e.g., advertisements, other people, the media, fiction, or ownership). Based on these measures, respondents of both samples were grouped according to their prior experience with the assigned VBA: (1) those who did not know the respective VBA before the study were categorized as having *no prior experience*, (2) those who knew the VBA before the study but only through indirect contact (e.g., advertisements, other people, the media, or fiction) were categorized as having *indirect experience,* and (3) those who said they owned the VBA were categorized as having *direct experience.* Among student respondents, 16% had no prior experience with their assigned VBA, 64% had indirect experience, and 20% said they had direct experience. This was similar in the staff sample, where 21% of respondents had no prior experience, 64% had indirect experience, and 15% said they had direct experience. This means that in both samples, most respondents had no or indirect prior experience with the VBA they had to assess in the survey. This was anticipated due to the low level of adoption of VBAs in Germany at the time of the survey (Newman et al., 2018; Taş et al., 2019), which is why the demonstrational survey design was employed. For these respondents, the assessment likely represents a first impression, and findings must be interpreted accordingly.

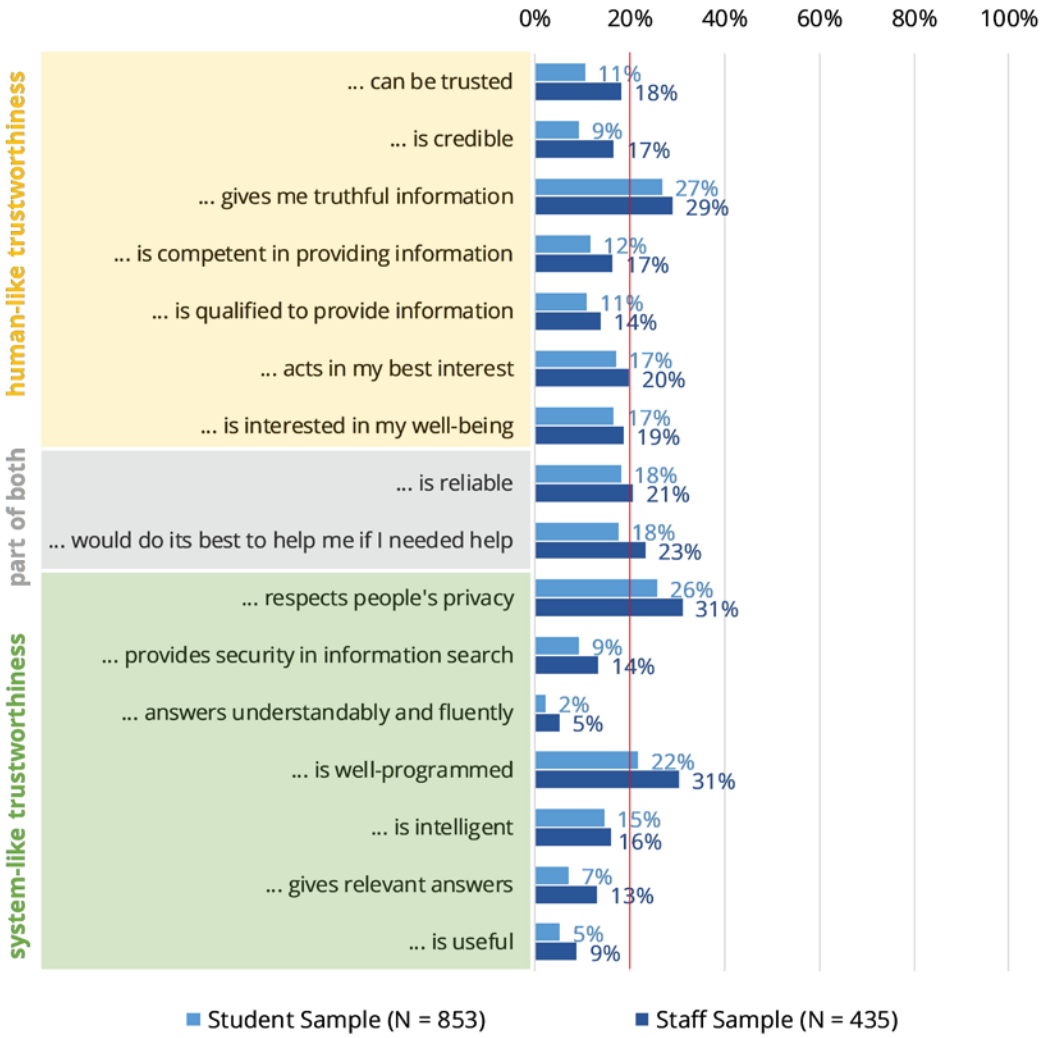
Results

This section is divided into four stages of inquiry. First, the percentages of the "I cannot assess that" responses to all individual trustworthiness items were examined and interpreted as a first indication of respondents' ability to apply the proposed items to VBAs in the context of news presentation. Second, to explore how a hybrid trustworthiness model for VBAs might look, an exploratory factor analysis (EFA) was conducted in the student sample with all 16 trustworthiness items from both the human-like and machine-like trustworthiness dimensions. The resulting model was then submitted to a confirmatory factor analysis (CFA) in the staff sample, to cross-validate it with this different cohort. Third, the applicability of the competing models—human, hybrid, or machine **(RQ1)**—was compared by conducting CFAs and contrasting the resulting model fit statistics. This procedure was applied in both samples to account for possible cohort effects. Last, the model comparison was repeated for different subgroups of respondents, which were specified according to the respondents' prior experience (i.e., none, indirect, or direct) with the VBA **(RQ2)**. Supplemental materials are available in this study's OSF repository.

Applicability of Individual Trustworthiness Items

As is apparent from Figure 1, most respondents from both samples were able to apply most of the items to VBAs. Student respondents had difficulties applying the human-like trustworthiness item *gives me truthful information*, as well as the two machine-like items *respects people's privacy* and is *well-programmed*. All of these may be caused by a lack of prior experience with the VBA in the context of news presentation, or a lack of contextual information available from the simulated interaction within the survey (Chita-Tegmark et al., 2021). Staff respondents had difficulties applying the same items but also struggled to assess the human-like item *acts in my best interest*, as well as both items that are conceptually part of both the human- and machine-like trustworthiness model: *is reliable* and *would do its best to help me if I needed help*. While the difficulty of assessing reliability may again result from a lack of experience or lack of contextual information, staff respondents' difficulties with the other two items may have been caused by a lack of perceived agency (Chita-Tegmark et al., 2021). Staff respondents may simply not perceive VBAs as having the ability to act, and, thus, they might not think that VBAs can help them.

FIGURE 1 **Percentage of "I cannot assess that" Responses to VBA Trustworthiness Items**



Note. Those items with percentages over 20% (red line) are discussed in detail.

In summary, more machine-like than human-like trustworthiness items caused assessment problems in both samples, which may be a first hint that they are not applicable to VBAs in the context of news presentation. Additional statistics (e.g., means, standard deviations, skewness, and kurtosis) are provided in the supplemental material.

Exploring the Hybrid Trustworthiness Model

To explore how a hybrid trustworthiness model including both human-like and machine-like items might look for VBA trustworthiness, an EFA was conducted with the student sample. The resulting model was then subjected to CFA in the staff sample to cross-validate it.

Student Sample: Exploratory Factor Analysis

First, the correlation matrix (supplemental material) for all items was examined. Thus, two items had to be excluded due to their low correlation (< 0.3) with more than one third of the other items, which suggests that they "do not 'fit' with the pool of items" (Field, 2018, p. 806). This concerned two machine-like items: the reliability item *respects people's privacy*, and the functionality item *answers understandably and fluently*. Both were excluded from the following EFA. Bartlett's test of sphericity ($\chi^2(91) = 2967.04$, $p < .001$), as well as the Kaiser-Meyer-Olkin measure (KMO = 0.935), confirmed the sampling adequacy for the analysis (Field, 2018). Because the literature suggests that the dimensions of trustworthiness correlate and interact with one another (Mayer et al., 1995), oblique rotation (promax, kappa = 4) was chosen to form factors. Items with communalities (after extraction) lower than 0.35 were excluded because low communalities suggest that the extracted factors only explain a small portion of variance in these items, sharing only a low level of common variance with the rest of the items (Field, 2018). This applied to the two machine-like functionality items *is well-programmed* and *is intelligent*.

The subsequent EFA with the remaining 12 items resulted in a two-factor solution (Table 2), according to Kaiser's criterion of eigenvalues over 1. Because machine-like reliability and helpfulness items, as well as human-like integrity and competence items, were clustered together on the first factor, this factor can be interpreted as *hybrid trustworthiness*, a true mix of both human and machine trustworthiness attributes. The second factor contained the three human-like *benevolence* items. Thus, for the student respondents, benevolence seems to stand out from the rest of the trustworthiness items when assessing VBAs, which may suggest that this dimension is inadequate for the trustworthiness assessment of these technologies. However, correlation between the two factors is high, with $r = .65$. Together, the two factors explained 62.71% of the variance.

Because of the established three-dimensional structure of previous trustworthiness models, the EFA was conducted again, this time forcing three factors to be extracted. As a result, the third extracted factor did not meet Kaiser's criterion, with an eigenvalue of 0.79. This eigenvalue is acceptable according to Jolliffe's more liberal criterion (Field, 2018). Therefore, the three-factor solution was maintained due to its agreement with previous trustworthiness definitions. Based on the way items were clustered together (Table 3), the three factors were interpreted as (1) *hybrid competence*, (2) *hybrid integrity*, and (3) *benevolence*—thus resembling dimensions found mostly in human trust literature, even if attributes from both human and machine trustworthiness were mixed. Together, the three factors explained 69.31% of the variance. Additionally, this solution improved the average communality after extraction. This factor solution is retained as an alternative hybrid trustworthiness model.

In summary, the EFAs in the student sample produced two alternative hybrid models for VBA trustworthiness.

TABLE 2 Hybrid Trustworthiness Model: 2-Factor Solution From EFA With the Student Sample

	Factor Loadings		
	Hybrid Trustworthiness	Benevolence	Communalities
... is credible	**0.86**	0.00	0.75
... is competent in providing information	**0.84**	-0.07	0.63
... gives me truthful information	**0.77**	-0.07	0.53
... is qualified to provide information	**0.73**	0.07	0.61
... is reliable	**0.71**	0.02	0.53
... provides security in information search	**0.70**	0.03	0.52
... can be trusted	**0.67**	0.13	0.59
... is useful	**0.67**	0.01	0.45
... gives relevant answers	**0.62**	0.05	0.42
... is interested in my well-being	-0.10	**0.76**	0.49
... would do its best to help me if I needed help	0.03	**0.73**	0.57
... acts in my best interest	0.14	**0.63**	0.54
Eigenvalues	6.31	1.12	6.63
% of variance	52.60	10.11	
Cronbach's α	0.91	0.78	

Notes. n = 443; KMO = .927; R^2 = 62.71%
Extraction Method: Principal Axis Factoring; Rotation Method: Promax with Kaiser Normalization
Factor loadings > 0,40 appear in bold

TABLE 3 Hybrid Trustworthiness Model: 3-Factor Solution From EFA With the Student Sample

	Factor Loadings			
	Hybrid Competence	Hybrid Integrity	Benevolence	Communalities
... is competent in providing information	**0.85**	0.06	-0.08	0.72
... is reliable	**0.80**	−0.04	0.02	0.62
... gives relevant answers	**0.55**	0.10	0.05	0.45
... is useful	**0.55**	0.15	0.01	0.47
... is qualified to provide information	**0.42**	0.35	0.07	0.60
... can be trusted	-0.07	**0.81**	0.12	0.69
... is credible	0.13	**0.81**	-0.02	0.82
... gives me truthful information	0.10	**0.74**	-0.09	0.59
... provides security in information search	0.27	**0.47**	0.03	0.52
... would do its best to help me if I needed help	0.21	−0.19	**0.76**	0.62
... is interested in my well-being	−0.17	0.09	**0.74**	0.49
... acts in my best interest	0.03	0.15	**0.61**	0.53
Eigenvalues	6.31	1.12	0.79	7.12
% of variance	52.60	10.11	6.61	
Cronbach's α	0.84	0.87	0.78	

Notes. n = 443; KMO = .927; R^2 = 69.31%
Extraction Method: Principal Axis Factoring; Rotation Method: Promax with Kaiser Normalization
Factor loadings > 0,40 appear in bold

Staff Sample: Confirmatory Factor Analysis

To cross-validate the two alternative hybrid trustworthiness models with the staff sample, confirmatory factor analyses (CFAs) were conducted using R (packages lavaan and semPlot). Both models were specified according to the findings from the EFAs in the student sample. The CFAs were interpreted by looking at a combination of recommended model fit statistics such as the comparative fit index (CFI > .95 for good fit), Tucker Lewis index (TLI > .95 for good fit), root mean square error of approximation (RMSEA < .08 for adequate fit), and standardized root mean square of residuals (SRMR < .06 for good fit) (Hu & Bentler, 1999; Kline, 2011; Pertegal et al., 2019; Weiber & Mühlhaus, 2014).

Based on the recommended thresholds for these fit statistics, the results of the two CFAs show that only the three-factor hybrid model exhibited an acceptable model fit in the staff sample, while the two-factor hybrid model did not meet the requirements for fit statistics (Table 4).[3] Thus, the two-factor hybrid model was rejected and only the three-factor hybrid model will be considered as the *hybrid model* in the following analyses.

TABLE 4 Hybrid Trustworthiness Model: Summary of Model Fit Statistics From CFAs With the Staff Sample

	Staff Sample (N = 435)	
	hybrid 2 factors	hybrid 3 factors
N	207	207
χ^2	221.00	102.99
df	53	51
GFI	0.82	0.93
CFI	0.88	0.96
TLI	0.85	0.95
RMSEA	0.12	0.07
RMSEA LL 90% CI	0.11	0.05
RMSEA UL 90% CI	0.14	0.09
SRMR	0.06	0.05

Notes. Acceptable fit statistics are highlighted in light green. Not acceptable fit statistics appear in red.
GFI – goodness of fit index >.90; CFI – comparative fit index >.95; TLI – Tucker-Lewis index >.95; RMSEA – root mean square error of approximation <.08; SRMR – standardized root mean square residual <.06

3. The resulting standardized factor loadings from the CFAs are available in this study's OSF repository.

Applicability of Different Trustworthiness Models: Human, Hybrid, or Machine?

To compare the human(-like), machine(-like), and hybrid models of VBA trustworthiness, additional CFAs were conducted in both samples to validate and compare the findings of this pilot study between two slightly different cohorts. While the exclusively human and exclusively machine models were specified based on the respective theoretical trustworthiness definitions, the hybrid model was specified according to the three-factor model that was exploratively developed with the student sample and validated with the staff sample. The applicability of the competing models was evaluated and contrasted by looking at the same combination of recommended model fit statistics as before. Thus, a model is interpreted as *applicable* if the results of the CFA indicate that its model fit statistics are above (goodness of fit) or below (badness of fit) the recommended thresholds. The model that exhibits the best model fit is interpreted as *most applicable* (i.e., to work best).

In both samples, the human-like, as well as the hybrid trustworthiness model, demonstrated an acceptable model fit. Meanwhile, in both samples, the machine-like model did not meet the requirements for fit statistics. Comparing the human-like and the hybrid trustworthiness model showed that the human-like model exhibited an overall better model fit according to conventional fit statistics. These results were uncovered using the student sample (Table 5), and they were validated using the staff sample (Table 6). Thus, the findings apply for two different cohorts. This suggests that a human-like model is indeed suitable for investigating VBAs' trustworthiness in the context of news presentation ***(*RQ1*)***.

TABLE 5 Model Comparison (RQ1): Summary of Model Fit Statistics From CFAs With the Student Sample

	Student Sample (N = 853)		
	human-like	**machine-like**	**hybrid**
N	459	413	443
χ^2	86.16	101.70	169.16
df	24	24	51
GFI	0.96	0.94	0.94
CFI	0.97	0.93	0.96
TLI	0.96	0.89	0.95
RMSEA	0.08	0.09	0.07
RMSEA LL 90% CI	0.06	0.07	0.06
RMSEA UL 90% CI	0.09	0.11	0.08
SRMR	0.04	0.05	0.04

Notes. Acceptable fit statistics are highlighted in light green. Best fit statistics are highlighted in dark green. Not acceptable fit statistics appear in red.
GFI – goodness of fit index >.90; CFI – comparative fit index >.95; TLI – Tucker-Lewis index >.95; RMSEA – root mean square error of approximation <.08; SRMR – standardized root mean square residual <.06

TABLE 6 Model Comparison (RQ1): Summary of Model Fit Statistics From CFAs With the Staff Sample

	Staff Sample (N = 435)		
	human-like	machine-like	hybrid
N	218	185	207
χ^2	38.53	67.07	102.99
df	24	24	51
GFI	0.96	0.92	0.93
CFI	0.99	0.90	0.96
TLI	0.98	0.85	0.95
RMSEA	0.05	0.10	0.07
RMSEA LL 90% CI	0.02	0.07	0.05
RMSEA UL 90% CI	0.08	0.13	0.09
SRMR	0.03	0.06	0.05

Notes. Acceptable fit statistics are highlighted in light green. Best fit statistics are highlighted in dark green. Not acceptable fit statistics appear in red.
GFI – goodness of fit index >.90; CFI – comparative fit index >.95; TLI – Tucker-Lewis index >.95; RMSEA – root mean square error of approximation <.08; SRMR – standardized root mean square residual <.06

The Moderating Role of Prior Experience

Finally, the competing models were compared regarding their fit according to the respondents' prior experience with the VBA. Model fit statistics for all comparisons described in this section are provided in the study's OSF repository. An overview of the applicability of all models according to the level of respondents' prior experience with the VBA is displayed in Table 7.

TABLE 7 Overview of Model Fit From CFAs According to Respondents' Prior Experience (RQ2)

Sample	Prior Experience	human-like	machine-like	hybrid
Student	None (n = 136)			x
	Indirect (n = 546)			
	Direct (n = 171)			
Staff	None (n = 90)		x	x
	Indirect (n = 280)		x	
	Direct (n = 65)	x	x	x

Notes. Models with acceptable model fit are highlighted in light green. The model exhibiting the best model fit for each subgroup is highlighted in dark green. Not acceptable models according to fit statistics are marked by a red x.

For students who had no prior experience with the VBA before the study, the machine-like model was most applicable (i.e., exhibited the overall best model fit), while the human-like model also exhibited an acceptable model fit. For those with indirect experience, the human-like model was most applicable, but both the machine-like and the hybrid model showed an acceptable fit as well. For those who owned the VBA themselves and thus have direct experience with the VBA, the human-like trustworthiness model still exhibited the best model fit but slightly less than for those with indirect experience. The model fit of the machine-like and hybrid model did not change and was also acceptable for this subgroup.

In the staff sample, the human-like model worked best for those who did not know the VBA before the study, as well as those with indirect experience. For the subgroup with indirect experience, the hybrid model also showed an acceptable fit. However, no model exhibited an acceptable model fit for those who had direct experience with the VBA, which may be due to the small size of this subgroup in the staff sample ($n = 65$).

Nonetheless, these results indicate that experience—even indirect experience—affects which models are suitable to assess communicative technologies (**RQ2**), as has been proposed by other scholars (e.g., Gambino et al., 2020; Horstmann & Krämer, 2019). Because this study was conducted in an early stage of VBA adoption in Germany, even those with direct experience might be relatively inexperienced. Over time, they may adapt their models further, and the hybrid model may be more applicable still.

Discussion of Results and Contributions

This study is a pilot study that takes a first step toward understanding how we can adequately investigate trustworthiness in emerging technologies that are perceptually hybrid and take on roles that were previously inherent to humans (e.g., communicator, news anchor). Due to their human-like CUI, VBAs are often attributed a mixture of human and machine characteristics, making them perceptual hybrids on the borderline between human(-like) and machine(-like). The paper discussed the implications of this perceptual hybridity of VBAs for the assessment of trustworthiness. As a result, both human-like and machine-like trustworthiness dimensions were considered to explore how a hybrid trustworthiness model could look.

In this explorative study, two well-interpretable hybrid trustworthiness models as a mix of both human-like and machine-like characteristics were discovered, one of which—the hybrid model with three factors—showed an acceptable model fit. However, when comparing model fit for the human, hybrid, and machine models, the *human* model exhibited the overall best model fit in both samples, which speaks in favor of the validity of this finding across different cohorts. This indicates that scales developed for human-human interactions might be, after all, applicable to VBAs.

The study also explored whether the applicability of the models differed between respondents with different levels of prior experience with the VBA (no prior experience, indirect experience, or direct experience). For those with no prior experience, this differed substantially between the two samples: While for students, the *machine* model exhibited the best model fit, for staff respondents, only the *human* model had an acceptable model fit. It seems that when people encounter VBAs for the first time (in an online survey, no less), which models are applicable, and thus how respondents interpret the available cues, depends largely on their predispositions.

When respondents of both samples had at least indirect experiences with the VBA, the human model exhibited the best model fit. Whether this is because the available secondhand information (e.g., advertisements or friends who just bought the VBA) emphasizes the human-likeness of VBAs and "triggers" this model is beyond the scope of this paper. However, the hybrid model also exhibited acceptable model fit at this level of prior experience, which may indicate that even secondhand information affects how hybrid communicators are assessed.

For students with direct experience with the VBA, the applicability of the human model decreased while that of the machine and hybrid model remained the same, thereby narrowing the fit gap between the competing models. Thus, it seems to depend on people's level of prior experience as to which characteristics they apply to assess trustworthiness when interacting with perceptually hybrid communicative technologies such as VBAs. Moreover, with increasing experience of the respondents, the applicability of the different models changed and started to match the perceptual hybridity of VBAs found in previous studies. These findings support the assumption that, as these hybrid technologies are increasingly embedded in our daily lives and routines, people's interaction scripts and mental models will evolve (e.g., Gambino et al., 2020; Horstmann & Krämer, 2019).

By testing alternative hybrid and non-hybrid models for VBA trustworthiness, this pilot study contributes to the ongoing discourse about adequate instruments for investigating communicative technologies. As is increasingly apparent, neither scales from social psychology developed for human-human interactions, nor those with a focus on usability or functionality, adequately recognize the nature of communicative technologies such as VBAs. Instead, they either over- or underestimate VBAs' perceptual human-likeness. While exploratory, this study demonstrated how the two previously distinct human and machine models of trustworthiness mix when the investigated trustee's nature blurs the distinction these models are based on (human-like or machine-like). Thus, this study provides further indications that previous scales may need to be adapted and new measures developed due to this perceptual hybridity. Future studies should apply a qualitative survey approach to explore whether there are additional, genuinely hybrid trustworthiness characteristics currently missing from the hybrid model.

Limitations and Directions for Future Research

The reported findings must be interpreted considering some limitations of the study. To achieve high internal validity for this pilot study and a large sample size, the data were collected via online surveys. Therefore, interactions with the VBAs had to be simulated by presenting prerecorded videos. It is possible that respondents may assess VBAs differently when directly interacting with them. Furthermore, this study was part of a larger project. In the beginning of both surveys, four additional videos were presented that contained the VBAs' answers to "personal" questions. These videos may have affected respondents' perceptions of the VBAs as more human-like, thus also possibly affecting the model fit of the human-like trustworthiness model. Hence, the hybrid model may have exhibited the best model fit if these videos had not been part of the survey. The connection between perception and assessment should be further investigated, for example, by exploring whether people apply different trustworthiness models according to how they ontologically classify VBAs (i.e., human, machine, or something in between) and how they perceive the VBAs' role in the communication process (i.e., source or channel).

This study found indications for the widely proposed adaptation of mental models for communicative technologies with growing experience. However, both samples were very similar and mostly knew the VBAs from indirect contact, the results may be most conclusive for initial trustworthiness assessments. However, because the study used real VBAs that are available on the market, two out of three respondents had at least indirect experience with the assigned VBA, and one in five student (and one in seven staff) respondents had direct experience with the assigned VBA, which is not the case for most laboratory experiments using fictious assistants or chatbots. Future studies should investigate whether the findings also apply to a more experienced population. Thus, environmental factors, such as perceived competence of the programmers or developer companies of a VBA, should be included as well to explore how these factors interact with the trustworthiness assessment of the VBA itself and how this affects the development of trust. Furthermore, VBA assessment should be tested with older participants and participants with a more diverse educational background.

In terms of context, different levels of perceived risks and expectations may be involved when VBAs present different kinds of information. Thus, future studies should investigate the applicability of different trustworthiness models for different types of information (e.g., news or service). Additionally, a media trustworthiness model should also be considered. A comparison of different versions of the same VBA (e.g., text-based, voice-based, and voice-based plus display) might also yield important insights into how the trustworthiness assessment of VBAs in the context of news presentation is affected by presentation mode.

Conclusion

This study explored how the perceptual hybridity of communicative technologies entails the need for adequate hybrid measurement models. This was tested for a construct that emphasizes distinguishing between humans and machines but is nonetheless important for both: trustworthiness. This pilot study contributes to that by examining whether previously distinct models of human and machine trustworthiness can be applied to VBAs, or whether their combination in a hybrid trustworthiness model better captures the perceptual hybridity of VBAs. By examining the applicability of the competing human, machine, and hybrid trustworthiness models in the context of news presentation, the study provides insights into people's mental models when assessing these technologies. While overall the human model had the best model fit in this study, the findings also supported a hybrid trustworthiness model. Additionally, model applicability was found to change with different levels of prior experience. With no prior experience, either the human or machine model worked best, depending on which cohort the respondents belonged to. When respondents had at least some (either indirect or direct) experience with the technology, the human, machine, and hybrid models were simultaneously applicable, suggesting that respondents had already begun to adapt their mental models. Thus, this paper argues that there is a need to develop hybrid measures that adequately recognize the hybrid nature of communicative technologies such as VBAs. Future research is needed to explore if there are additional, genuinely hybrid trustworthiness characteristics, and whether they supplement or replace human and machine trustworthiness characteristics.

Author Biography

Lisa Weidmüller is a PhD student and research associate at the Institute of Media and Communication at TU Dresden. Her research focuses on interaction with and through emerging digital technologies and information intermediaries, especially regarding acceptance and development of trust.

https://orcid.org/0000-0003-1680-9122

Center for Open Science

OPEN DATA OPEN MATERIALS

This article has earned the Center for Open Science badges for Open Data and Open Materials. Stimulus, data, and analysis files are available in this project's supplemental materials: https://osf.io/rnq39.

References

Akter, S., D'Ambra, J., & Ray, P. (2011). Trustworthiness in mHealth information services: An assessment of a hierarchical model with mediating and moderating effects using partial least squares (PLS). *Journal of the American Society for Information Science and Technology*, *62*(1), 100–116. https://doi.org/10.1002/asi.21442

Ammari, T., Kaye, J., Tsai, J. Y., & Bentley, F. (2019). Music, search, and IoT: How people (really) use voice assistants. *ACM Transactions on Computer-Human Interaction*, *26*(3), 1–28. https://doi.org/10.1145/3311956

Blöbaum, B. (2016). Key factors in the process of trust. On the analysis of trust under digital conditions. In B. Blöbaum (Ed.), *Trust and communication in a digitized world: Models and concepts of trust research* (pp. 3–25). Springer International Publishing. https://www.springer.com/de/book/9783319280578

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Ramirez, A., Dunbar, N. E., & Miczo, N. (1999). Testing the interactivity model: Communication processes, partner assessments, and the quality of collaborative work. *Journal of Management Information Systems*, *16*(3), 33–56. https://doi.org/10.1080/07421222.1999.11518255

Chita-Tegmark, M., Law, T., Rabb, N., & Scheutz, M. (2021). Can you trust your trust measure? *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 92–100. https://doi.org/10.1145/3434073.3444677

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., Edwards, J., & Cowan, B. R. (2019). The state of speech in HCI: Trends, themes, and challenges. *Interacting with Computers*, *31*(4), 349–371. https://doi.org/10.1093/iwc/iwz016

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, *92*(4), 909–927. https://doi.org/10.1037/0021-9010.92.4.909

Deloitte. (2018). *Beyond touch: Voice-commerce 2030*. https://web.archive.org/web/20211228170629/https://www.deloitte-mail.de/custloads/141631293/Deloitte-Beyond-Touch-Studie.pdf

Easwara Moorthy, A., & Vu, K.-P. L. (2015). Privacy concerns for use of voice activated personal assistant in the public space. *International Journal of Human-Computer Interaction*, *31*(4), 307–335. https://doi.org/10.1080/10447318.2014.986642

Edwards, A. P. (2018). Animals, humans, and machines: Interactive implications of ontological classification. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 29–50). https://doi.org/10.3726/b14399

Edwards, A. P., Edwards, C., Westerman, D., & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior*, *90*, 308–314. https://doi.org/10.1016/j.chb.2018.08.042

Edwards, J., & Sanoubari, E. (2019). A need for trust in conversational interface research. *Proceedings of the 1st International Conference on Conversational User Interfaces*, 1–3. https://doi.org/10.1145/3342775.3342809

Etzrodt, K., & Engesser, S. (2021). Voice-based agents as personified things: Assimilation and accommodation as equilibration of doubt. *Human-Machine Communication*, *2*, 57–79. https://doi.org/10.30658/hmc.2.3

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Fink, D. (2014). *Vertrauen in autonome Technik [Trust in autonomous technology]* [TU Dortmund]. http://dx.doi.org/10.17877/DE290R-15509

Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*(1). https://doi.org/10.30658/hmc.1.5

Gao, L., & Bai, X. (2014). A unified perspective on the factors influencing consumer acceptance of internet of things technology. *Asia Pacific Journal of Marketing and Logistics*, *26*(2), 211–231. https://doi.org/10.1108/APJML-06-2013-0061

Garcia, M. P., Lopez, S. S., & Donis, H. (2018, July 1). Voice activated virtual assistants personality perceptions and desires: Comparing personality evaluation frameworks. *Proceedings of British HCI 2018*. 32nd International BCS Human Computer Interaction Conference, Belfast, UK. https://doi.org/10.14236/ewic/HCI2018.40

Gunkel, D. J. (2020). *An introduction to communication and artificial intelligence*. Polity Press.

Guzman, A. L. (2015). *Imagining the voice in the machine: The ontology of digital social agents*. University of Illinois at Chicago.

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, *90*, 343–350. https://doi.org/10.1016/j.chb.2018.08.009

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for Human-Machine Communication. *Human-Machine Communication*, *1*(1). https://doi.org/10.30658/hmc.1.3

Hearst, M. A. (2011). "Natural" search user interfaces. *Communications of the ACM*, *54*(11), 60–67. https://doi.org/10.1145/2018396.2018414

Horstmann, A. C., & Krämer, N. C. (2019). Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00939

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion. Psychological studies of opinion change*. Yale University Press. http://psycnet.apa.org/record/1954-05952-000

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kahn, Jr., P. H., Reichert, A. L., Gary, H., Kanda, T., Ishiguro, H., Shen, S., Ruckert, J. H., & Gill, B. (2011). The new ontological category hypothesis in human-robot interaction. *Proceedings of the 6th International Conference on Human Robot Interaction (HRI)*, 159–160. https://doi.org/10.1145/1957656.1957710

Kinsella, B., & Mutchler, A. (2020). *Smart speaker consumer adoption report*. voicebot.ai. https://research.voicebot.ai/report-list/smart-speaker-consumer-adoption-report-2020/

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed). Guilford Press.

Koh, Y. J., & Sundar, S. S. (2010). Effects of specialization in computers, web sites, and web agents on e-commerce trust. *International Journal of Human-Computer Studies*, *68*(12), 899–912. https://doi.org/10.1016/j.ijhcs.2010.08.002

Lankton, N., McKnight, D. H., & Thatcher, J. B. (2014). Incorporating trust-in-technology into Expectation Disconfirmation Theory. *The Journal of Strategic Information Systems*, *23*(2), 128–145. https://doi.org/10.1016/j.jsis.2013.09.001

Lankton, N., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 880–918. https://doi.org/10.17705/1jais.00411

Lee, C.-J., Scheufele, D. A., & Lewenstein, B. V. (2005). Public attitudes toward emerging technologies: Examining the interactive effects of cognitions and affect on public attitudes toward nanotechnology. *Science Communication*, *27*(2), 240–267. https://doi.org/10.1177/1075547005281474

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

Lei, X., Tu, G.-H., Liu, A. X., Li, C.-Y., & Xie, T. (2018). The insecurity of home digital voice assistants—Vulnerabilities, attacks and countermeasures. *IEEE Conference on Communications and Network Security (CNS)*, 1–9. https://doi.org/10.1109/CNS.2018.8433167

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, *13*(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403

Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems*, *17*(1), 39–71. https://doi.org/10.1016/j.jsis.2008.01.001

Lombard, M., & Xu, K. (2021). Social responses to media technologies in the 21st century: The Media Are Social Actors paradigm. *Human-Machine Communication*, *2*, 29–55. https://doi.org/10.30658/hmc.2.2

Lyons, K. (2020, April 16). *Amazon's Alexa gets a new longform speaking style*. The Verge. https://web.archive.org/web/20200416230934/https://www.theverge.com/2020/4/16/21224141/amazon-alexa-long-form-speaking-polly-ai-voice

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709. https://doi.org/10.2307/258792

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 1–25. https://doi.org/10.1145/1985347.1985353

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

McTear, M., Callejas, Z., & Barres, D. G. (2016). *The conversational interface: Talking to smart devices*. Springer International Publishing. https://doi.org/10.1007/978-3-319-32967-3

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & Mccann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Annals of the International Communication Association*, *27*(1), 293–335. https://doi.org/10.1080/23808985.2003.11679029

Moon, Y., Kim, K. J., & Shin, D.-H. (2016). Voices of the internet of things: An exploration of multiple voice effects in smart homes. In N. Streitz & P. Markopoulos (Eds.), *Distributed, ambient and pervasive interactions* (pp. 270–278). Springer International Publishing. https://doi.org/10.1007/978-3-319-39862-4_25

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Natale, S. (2021). To believe in Siri: A critical analysis of voice assistants. In *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press. https://doi.org/10.1093/oso/9780190080365.003.0007

Natale, S., & Cooke, H. (2020). Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces. *Media, Culture & Society*, 0163443720983295. https://doi.org/10.1177/0163443720983295

Newhagen, J., & Nass, C. (1989). Differential criteria for evaluating credibility of newspapers and TV news. *Journalism Quarterly*, *66*(2), 277–284. https://doi.org/10.1177/107769908906600202

Newman, N. (2018). *The future of voice and the implications for news*. Reuters Institute for the Study of Journalism. https://web.archive.org/web/20211228171333/https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-11/Newman%20-%20Future%20of%20Voice%20FINAL_1.pdf

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2018). *Digital news report 2018*. Reuters Institute for the Study of Journalism. https://web.archive.org/web/20190520173020/http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39–52. https://doi.org/10.1080/00913367.1990.10673191

Olson, C., & Kemery, K. (2019). *Microsoft voice report*. https://web.archive.org/web/20190502032155/https://about.ads.microsoft.com/en-us/insights/2019-voice-report

Oxford University Press. (n.d.). Hybrid. In *OED Online*. https://web.archive.org/web/20211228171746/https://www.oed.com/view/Entry/89809

Pertegal, M.-Á., Oliva, A., & Rodríguez-Meirinhos, A. (2019). Development and validation of the Scale of Motives for Using Social Networking Sites (SMU-SNS) for adolescents and youths. *PLOS ONE*, *14*(12), e0225781. https://doi.org/10.1371/journal.pone.0225781

Porter, J. (2019, January 16). *Alexa's news-reading voice just got a lot more professional*. The Verge. https://web.archive.org/web/20190116184000/https://www.theverge.com/2019/1/16/18185258/alexa-news-voice-newscaster-news-anchor-us-launch

Priest, S. H. (2001). Misplaced faith: Communication variables as predictors of encouragement for biotechnology development. *Science Communication*, *23*(2), 97–110. https://doi.org/10.1177/1075547001023002002

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Introduction to special topic forum: Not so different after all: A cross-discipline view of trust. *The Academy of Management Review*, *23*(3), 393–404.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Statista. (2021). *Global smart speaker vendors' market share 2016–2020*. https://web.archive.org/web/20210124041320/https://de.statista.com/statistik/daten/studie/818999/umfrage/marktanteile-der-hersteller-am-absatz-von-intelligenten-lautsprechern-weltweit/

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction. *Communication Research*, *27*(6), 683–703. https://doi.org/10.1177/009365000027006001

Taş, S., Hildebrandt, C., & Arnold, R. (2019). *Sprachassistenten in Deutschland [Voice assistants in Germany]* (Discussion paper No. 441). Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste. https://web.archive.org/web/20211228172352/https://www.wik.org/uploads/media/WIK_Diskussionsbeitrag_Nr_441.pdf

Ullman, D., & Malle, B. F. (2018). What does it mean to trust a robot? Steps toward a multidimensional measure of trust. *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 263–264. https://doi.org/10.1145/3173386.3176991

Wang, W., & Benbasat, I. (2016). Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents. *Journal of Management Information Systems*, *33*(3), 744–775. https://doi.org/10.1080/07421222.2016.1243949

Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung [Structural equation models]*. Springer-Verlag. https://doi.org/10.1007/978-3-642-35012-2

Yang, X., Aurisicchio, M., & Baxter, W. (2019). Understanding affective experiences with conversational agents. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300772