Computers in Human Behavior 119 (2021) 106727

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Full length article

Dual humanness and trust in conversational AI: A person-centered approach

Peng Hu [a], Yaobin Lu [a,*], Yeming (Yale) Gong [b]

[a] School of Management, Huazhong University of Science & Technology, Luoyu Road, 1037, Wuhan, China
[b] Business Intelligence Center, EMLYON Business School, Yon, France

ARTICLE INFO

Keywords:
Artificial intelligence
Humanness perception
Trust
Person-centered approach
Finite mixture modeling

ABSTRACT

Conversational Artificial Intelligence (AI) is digital agents that interact with users by natural language. To advance the understanding of trust in conversational AI, this study focused on two humanness factors manifested by conversational AI: speaking and listening. First, we explored users' heterogeneous perception patterns based on the two humanness factors. Next, we examined how this heterogeneity relates to trust in conversational AI. A two-stage survey was conducted to collect data. Latent profile analysis revealed three distinct patterns: para-human perception, para-machine perception, and asymmetric perception. Finite mixture modeling demonstrated that the benefit of humanizing AI's voice for competence-related trust can evaporate once AI's language understanding is perceived as poor. Interestingly, the asymmetry between humanness perceptions in speaking and listening can impede morality-related trust. By adopting a person-centered approach to address the relationship between dual humanness and user trust, this study contributes to the literature on trust in conversational AI and the practice of trust-inducing AI design.

1. Introduction

Conversational artificial intelligence (AI) refers to digital agents that use natural language to communicate with users (Khatri et al., 2018). With the recent advances in natural language processing (NLP), conversational AI is becoming increasingly prevalent in daily life. On one hand, it has been integrated into several devices, such as smartphones (e.g., Apple's Siri and Google Assistant) and speakers (e.g., Amazon's Echo and Google Home). On the other hand, it has also been embedded in many contexts, such as in-vehicle assistants, call center chatbots, and hospital guidance robots (Gursoy, Chi, Lu, & Nunkoo, 2019). Conversational AI can produce a human-like voice and listen to users' words as a human would, thereby transforming the mode of human-computer interaction.

However, market studies have found that this improved interactive mode has not engendered trust. For example, Microsoft found that 41% of voice assistant users had concerns about trust and privacy (Olson, 2019). Another market research found that 73% of consumers were unlikely to trust an AI assistant to make simple telephone calls correctly (Martin, 2019). At the same time, AI companies have been on a journey to monetize the commercial value of conversational AI by developing voice shopping, which allows users to shop online by talking to a voice assistant, such as Amazon's Alexa, Google Assistant, and Alibaba's Tmall Genie (Klaus & Zaichkowsky, 2020; Rhee & Choi, 2020). Trust has been recognized as a crucial facilitator in commerce-related behavior, such as online shopping (Gefen, Karahanna, & Straub, 2003). Users are unlikely to shop via conversational AI when they do not trust their devices. Thus, the lack of trust in conversational AI can prevent companies from unlocking the business potential of this new technology. Enhancing user trust is a common challenge for scholars and managers.

Trust in technology is an important research topic in human-computer interaction (HCI) (Lankton, McKnight, & Tripp, 2015). However, the existing research on trust in conversational AI is not abundant. Although previous studies examined how users perceive various human-like characteristics of conversational AI (e.g., voice pitch, voice accent, voice gender), most of the focused characteristics have been limited to the speaking aspect of conversation AI (Chang, Lu, & Yang, 2018; Edwards, Edwards, Stoll, Lin, & Massey, 2019; Niculescu, van Dijk, Nijholt, Li, & See, 2013; Tamagawa, Watson, Kuo, MacDonald, & Broadbent, 2011). Few studies have considered the listening aspect of conversation AI when examing humanness perception, which however is also vital for natural language interaction. Besides communicating in a

* Corresponding author.
E-mail addresses: hpeng@hust.edu.cn (P. Hu), luyb@hust.edu.cn (Y. Lu), gong@em-lyon.com (Y.(Y. Gong).

https://doi.org/10.1016/j.chb.2021.106727
Received 5 September 2020; Received in revised form 15 December 2020; Accepted 29 January 2021
Available online 2 February 2021
0747-5632/

human-like voice, conversational AI should also be capable of understanding what users say in a human-like way. Therefore, it is important to address how users distinctly perceive the implied humanness in the speaking and listening aspects of conversational AI and its association with user trust.

Adopting an interactive perspective, this study proposes a dualistic model of humanness perception for conversational AI. The model comprises *voice humanness perception* reflecting perceived humanness regarding the speaking aspect of conversation AI, and *understanding humanness perception* reflecting perceived humanness concerning the listening aspect of conversation AI. Based on the dualistic model, the current study addresses the following research questions:

RQ1: Are there user groups who show heterogeneous perception patterns of conversational AI's humanness?

RQ2: If yes, how does the level of trust in conversational AI vary among users with heterogeneous perception patterns of conversational AI's humanness?

These research questions are answered by a person-centered study. Particularly, we conducted a two-stage online survey on the users of conversational AI. In the first stage, we collected a pilot sample to validate the perceived humanness measures, and in the second stage, we collected a formal sample to answer our research questions. Latent profile analysis was used to identify the unobserved perception heterogeneity as asked in RQ1, and finite mixture modeling was used to test the differences in user trust among the uncovered perception patterns as asked in RQ2. Finally, we performed robustness checks to justify the results with a set of additional analyses.

This study contributes to the humanness perception literature and design practices for trust-inducing AI. First, this study reveals three latent profiles of humanness perception for conversation AI that deepens our understanding of how users distinctly perceive the humanness conveyed by the speaking and listening of conversation AI (Lortie & Guitton, 2011). Second, the prior studies claim that integrating a human-like voice can improve user trust (Chang et al., 2018; Edwards et al., 2019; Niculescu et al., 2013; Tamagawa et al., 2011). But this study indicates that the benefit of some humanized voice design for competence-based trust may evaporate once the natural language understanding of conversation AI is perceived as poor. Therefore, this study extends our knowledge concerning the relationship between voice humanization and user trust. Third, this study finds that users with asymmetric perception patterns show the lowest level of benevolence and integrity-based trust. This finding is important because it suggests that the asymmetry between different humanness dimensions may undermine morality-related trust. Practically, our findings suggest that managers should be cautious about the decision to humanize AI devices' voice when the language understanding is perceived as poor, and be more attentive to the potential imbalance between the speaking and listening of AI devices.

2. Literature review

2.1. Human-like design of conversational AI

With the rapid development of natural language processing and machine learning in recent years, the issue of users' responses to the human-like attributes in conversational AI has received attention (Chang et al., 2018; Edwards et al., 2019; Niculescu et al., 2013; Tamagawa et al., 2011; Torre, Goslin, & White, 2020; Xu, 2019). Those examined human-like attributes can be classified into two types: verbal and verbal-related. For verbal cues, researchers investigate users' responses to diverse verbal features. For example, Tamagawa et al. (2011) show that robots with a local accent can receive more positive evaluation than robots with accents elsewhere. Niculescu et al. (2013) unravel that the voice pitch of social robots can also affect users' ratings, with higher-pitched ones being rated as more attractive. Xu (2019) compares users' social responses to human voice versus synthetic voice and uncovers that users are inclined to develop high trust in social bots with a human voice than with a synthetic voice. An obvious characteristic of this literature stream is focusing on verbal cues.

For verbal-related cues, researchers examine how humanness clues reflected by conversational AI's voice (e.g., gender, age, personality, and emotion) influences users' perceptions and evaluations. For instance, Chang et al. (2018) demonstrates that users prefer female and extroverted voice, which indeed have been deployed in many digital assistants or call centers. Edwards et al. (2019) find that higher age participants appraise the older AI voice instructor as more credible and social, supporting social identity theory in human-AI interaction. Torre et al. (2020) explore the effect of smiling in conversational agents' voice on users' trust, and find that a smiling voice can continuously increase trust even when untrustworthy evidence are presented to users. Overall, prior studies in the emerging field of human-AI interaction have addressed some important issues with valuable insights.

However, whether the examined objects were verbal cues or verbal-related cues in conversational AI, the extant research focused on how the *speaking* aspect of AI systems influences user attitude and behavior (Chang et al., 2018; Edwards et al., 2019; Niculescu et al., 2013; Tamagawa et al., 2011; Torre et al., 2020; Xu, 2019), while the *listening* aspect of AI systems received little attention. Speaking and listening are the two indispensable elements for natural language interaction. According to the communicative competence theory, the communicative competence of an interlocutor can be classified into two aspects: speaking competence and listening competence (Haas & Arnold, 1995). To achieve effective human-human communication, one should speak in an appropriate way that can be understood by others but also correctly understand what others say, and deficiency in any either side would reduce the communicative effectiveness (Duran, 1983). Similarly, for human-AI communication, conversational AI that only has a human-like voice cannot achieve natural language interaction with users, because users' natural language feedback must also be comprehended by conversational AI (Braun, Broy, Pfleging, & Alt, 2019; Foehr & Germelmann, 2020; Santos, Abreu, Beça, Rodrigues, & Fernandes, 2020). For a bidirectional and effective voice interaction, conversational AI should be able to both output human-like voice and understand users' natural language input, none of which can be absent (Foehr & Germelmann, 2020).

On the other hand, in human-human communication, an excellent speaker is not necessarily an excellent listener and vice versa. Likewise, this is extremely true for human-AI conversation because the underlying technologies supporting the speaking and listening of conversational AI are different. The former is natural speech synthesis, while the latter is natural language understanding (Braun et al., 2019). The difference in underlying technology implies that users may form divergent humanness perceptions regarding speaking and listening because of the potential technology-level imbalance among them (Nass & Moon, 2000), which further suggests that the two humanness perceptions should be concurrently considered when studying the humanness phenomenon in conversational AI.

Moreover, the listening aspect of conversation AI is important for user trust for two reasons. First, the listening ability is a central indicator of conversational AI's functional performance (Santos et al., 2020). If a conversational AI often misunderstands users' words in daily voice interaction, then users are unlikely to trust it in performing various tasks. Second, the interpersonal trust theory suggests that people are inclined to build trust relationships with others who can understand their words easily in daily conversation because they believe that this ease is caused by their shared values and that those people are easy to communicate with when involved in a cooperation activity (Gillath et al., 2020; Rotter, 1971). This may be also true for developing trust between users and AI systems as nowadays AI systems are becoming more and more human-like in various ways.

The interactivity of conversational AI not only challenges the existing knowledge about how users interact with systems (Schuetz &

Venkatesh, 2020), but also provides a unique opportunity to deepen the understanding of humanness perception in the context of conversation AI. Humanness perception of technology is defined as the degree to which a user feels a certain technology or system is human-like (versus machine-like) (Lankton et al., 2015; Schuetzler, Grimes, & Scott Giboney, 2020). In general, the level of perceived humanness is dependent on what human-related attributes a technology has (e.g., voice) and how intensively those attributes are manifested by the technology (e.g., synthetic voice vs. humanoid voice). Specifically, humanness perception is mainly indicated by the following three aspects of technology: social presence, social affordances, and affordances for sociality (Lankton et al., 2015). Social presence denotes the ability of technology to convey social cues, such as the smiling voice of conversational AI (Torre et al., 2020). Social affordances are technology-offered action potentials to a user via its social nature, such as the conversation action potential enabled by the NLP technology (Braun et al., 2019). Affordances for sociality are technology-offered action potentials enabling users to interact with others, such as using conversational AI to call friends (Khatri et al., 2018). If a technology possesses a high social presence, offer many social affordances or affordances for sociality, it can be perceived as human-like.

In the conversational AI field, previous studies have identified several factors that influence the humanness perception of conversational AI. For example, Gnewuch, Morana, Adam, and Maedche (2018) finds that chatbots using dynamic response delays can increase users' perceived humanness. Svenningsson and Faraon (2019) uncovers that natural flow of conversation, words choice, context-based emotional expression are related to perceived humanness in conversational agents. More recently, Schuetzler et al. (2020) examined two conversational skills of chatbots: response tailoring and response variety, and reveals that people perceive higher humanness in chatbots when chatbots give responses that are tailored to the current conversation or deploy different words to convey the same meaning throughout human-chatbot interaction. These studies improve our understanding of what characteristics of conversational AI can result in humanness perception.

However, the existing literature lacks a framework for organizing these detailed humanness characteristics regarding conversational AI. Inspired by communicative competence theory, this study takes an interactive perspective to propose a dualistic model of humanness perception for conversational AI: voice humanness perception (VHP) and understanding humanness perception (UHP). Based on the existing humanness perception literature (de Kleijn, Wijnen, & Poletiek, 2019; Lankton et al., 2015; Schuetzler et al., 2020), VHP is defined as the degree to which a user feels that the speaking aspect of a conversational AI system is human-like, and UHP is defined as the degree to which a user feels that the listening aspect of a conversational AI system is human-like (Sheehan, Jin, & Gottlieb, 2020; Westerman, Cross, & Lindmark, 2019).

VHP is primarily dependent on how a conversational AI speak using a natural tone, such as tailored responses based on conversation context and employ diverse words to convey a message (Schuetzler et al., 2020), while UHP is mainly dependent on how a conversational AI performs in understanding what users say, and only those conversational AI possessing good comprehension are likely to perceived as human-like regarding natural language understanding (Braun et al., 2019; Foehr & Germelmann, 2020; Santos et al., 2020). Analogous to human-human communication, voice humanness corresponds to the speaking attribute of conversational AI, while understanding humanness corresponds to the listening attribute of conversational AI (Wise & Hsiao, 2019). Thus, these two concepts are distinct from each other, but combining the two aspects would provide the potential for humans to interact with AI using natural language as in interpersonal communication.

2.2. Trust in conversational AI

Trust in technology is an important topic in the HCI field (Lankton et al., 2015; Xie, Prybutok, Peng, & Prybutok, 2020). As a new generation of technology, the trust in AI may be more complex than the trust in traditional technology because AI exhibits several human-like capabilities such as learning and reasoning (Gillath et al., 2020). How to enable trustworthy AI has become a common challenge for researchers and practitioners. Indeed, a recent study has unearthed the "algorithm aversion" phenomenon which shows that people are often unwilling to adopt AI algorithms even though these algorithms outperform humans in specific tasks (Dietvorst, Simmons, & Massey, 2016). These examined AI are algorithm-formed that cannot "talking to users". Does "algorithm aversion" still hold for AI that can communicate with users in natural language? In other words, how the unique elements of conversational AI – *speaking* and *listening* – associate with user trust? Although scholars have embarked on addressing the issue of trust in algorithmic AI, the extant research about understanding users' trust in conversational AI is scant.

With some exceptions, a design science study develops design guidelines for in-vehicle virtual assistants, in which they stress that in-vehicle virtual assistants should be designated with a consistent voice to foster users' trust (Strohmann, Siemon, & Robra-Bissantz, 2019). Another qualitative study finds that the perceived personality of technologies' voice acts as a crucial path for consumers building trust relationships with them (Foehr & Germelmann, 2020). The scope of the two studies above is still limited to the aspect of VHP in our proposed dualistic model, without considering UHP. Other studies investigate users' trust in conversational AI from the angle of privacy concerns or interaction quality (Ahmadian & Lee, 2017; Saffarizadeh, Boodraj, & Alashoor, 2017). These factors are important, but it is also necessary to understand how the unique factors of conversation AI shape user trust beyond those general factors, to advance the understanding of trust in conversation AI, which constitutes the original intention of the current study.

Trust is a multi-dimensional concept, and previous research indicates that the dimensionality of trust depends on the humanness level of the trusted technology. Particularly, for a less human-like technology (e.g., Microsoft Excel), users' trust is built on evaluations of the focal technology's reliability, functionality, and helpfulness. But for a more human-like technology (e.g., conversational AI in this study), users' trust is based on evaluations of the focal technology's integrity, competence, and benevolence (Lankton et al., 2015). The mismatch between trust components and technology humanness may confuse or mislead users in the trust evaluation process. Thus, in this study, we treat trust in conversational AI as a multifaceted construct comprising dimensions of integrity, competence, and benevolence, because the conversational AI possesses many human-like characteristics that traditional information systems do not have, such as human-like voice output and human-like understanding of users' voice input.

Inspired by Mayer, Davis, and Schoorman (1995) and McKnight, Choudhury, and Kacmar (2002), we provide the contextualized definitions for the three dimensions above. First, competence refers to the belief that a conversational AI can do what the user needs to have done, which is the ability-related trust. Second, integrity denotes the belief that a conversational AI sticks to a set of principles that the user can accept, while benevolence refers to the belief that a conversational AI will want to do good to the user apart from a profit motive, both of which are not ability-unrelated but moral-related trust (Wang & Benbasat, 2016). Indeed, morality is originally a concept in the realm of human behavior. But recently AI ethics has become a hot topic but also a tricky challenge. Thus, differentiating these dimensions and capturing both ability-related and moral-related trust in this study is of great significance to understand people's potentially differing confidence in conversational AI's functional ability and societal morality. Especially, this study deploys a person-centered approach to explore the heterogeneity of humanness perception patterns of conversational AI based on the proposed dualistic model and examine how the heterogeneity relates to users' trust in conversational AI, which will be detailed in the following.

2.3. Person-centered approach vs. variable-centered approach

The variable-centered approach is the traditional and dominant methodological view in social science, under which researchers focus on the relationships among a set of variables (Howard & Hoffman, 2018). Adopting this approach, the previous research has examined the relationships between several humanness cues of conversational AI and user experience (Chang et al., 2018; Edwards et al., 2019; Niculescu et al., 2013), and offers important insights into AI design and management. However, the fundamental assumption underlying this approach is that the sample, and the population from which it is extracted, are homogeneous. This assumption is difficult to be true, and misleading results can be gained once this prerequisite is unsatisfied (Meyer, Stanley, & Vandenberg, 2013).

The person-centered approach provides a crucial methodological angle to handle the potential heterogeneity. Instead of focusing on how variables relate to each other, this approach focuses on why individuals respond to these variables in diverse styles and how these heterogeneous response styles shape outcomes (Woo, Jebb, Tay, & Parrigon, 2018). In other words, the person-centered approach first relaxes the homogeneity assumption of the research sample and its population, and probe the existence of latent classes within which individuals share a similar pattern on a set of variables, but between which individuals have distinct patterns on those variables. This methodological angle treats "person" (not "variable") as the starting point. Although this approach can also be used to further examine the relationships among variables, the "variables" here are no longer the same meaning as that in the variable-centered approach because the kennel variable in person-centered research is an unobserved latent class, based on which researchers may seek the antecedents of the class membership and its influences on outcomes of interest (Howard & Hoffman, 2018).

The two approaches are both important for theory development. They represent two mindsets to view the phenomena of interest (Zyphur, 2009). Also, they can be complementary with each other to prompt theoretical progress. But for practical impacts, the person-centered approach may be more powerful because people, including managers, have an innate inclination to think in a categorized way (Zyphur, 2009). In sum, different from prior works, we take person as our outset to explore the existence of heterogeneous perception patterns in conversational AI users in terms of voice humanness perception and understanding humanness perception, and to further investigate the relationship between the unobserved heterogeneity and user's trust in conversational AI. The findings derived from this methodological lens can provide novel insights into the emerging issue of trust in conversational AI.

3. The rationale of this study

The theme of this study is to address trust in conversational AI from a person-centered viewpoint. Specifically, the objectives of the current study are twofold. First, we use latent profile analysis to examine the occurrence of users' unobserved heterogeneous perception patterns in terms of conversational AI's humanness, based on the proposed dualistic taxonomy (i.e., voice humanness perception and understanding humanness perception). Second, we deploy finite mixture modeling to examine the relationship between the heterogeneity of perception patterns and user's multifaceted trust in terms of competence, integrity, and benevolence of conversational AI.

As the first step to provide new insights regarding trust in conversational AI from a person-centered angle and the exploratory nature of this methodological paradigm (Meyer et al., 2013), we organize this paper in an inductive manner. Nevertheless, we have at least two rationales for the existence of the heterogeneous perception patterns. First, the user perception is subjective in nature, thereby differing humanness level could be perceived even with the same conversational AI, due to the various user characteristics mirroring individual differences. For example, users who have little prior experience with AI devices or applications may perceive one conversation AI as higher in humanness than that of users who have more prior experience such as AI trainers.

Second, conversation AI's voice humanness and understanding humanness are distinctive from each other. Voice humanness is supported by speech synthesis technology, while understanding humanness is achieved through speech recognition and natural language understanding technologies. Naturally, a conversational AI with high voice humanness is not necessarily to be high in understanding humanness, and vice versa. Moreover, based on extant evidence supporting the influence of technological humanness on user trust (de Visser et al., 2016; Go & Sundar, 2019; Qiu & Benbasat, 2009; Waytz, Heafner, & Epley, 2014), users' trust in conversation AI may be dependent on the specific perception patterns they belong to. Again, we do not develop specific hypotheses for the relations between the perception pattern of humanness and trust level. Instead, we utilize this opportunity to provide novel insights into AI research and practice regarding the topic of human-like design of AI systems and user trust.

4. Method

Fig. 1 illustrates the overall process of the current study.

4.1. Measures

Following recent research on trust in human-like IT artifacts (de Visser et al., 2016; Lankton et al., 2015; Waytz et al., 2014), we adopted the three-dimension view of trust construct — competence, benevolence, and integrity — and measure trust in conversational AI with a 10-item scale revised from McKnight et al. (2002). Specifically, the subscale of competence measures users' beliefs on the ability or performance of conversational AI with 4 items. An example item is "my voice assistant performs all of its roles very well." The subscale of benevolence assesses users' beliefs on how conversational AI would care for their interests with 3 items. An example item is "I believe that my voice assistant would act in my best interest." The subscale of integrity appraises users' beliefs on how AI would stick to a set of principles that users can accept with 3 items. An example item is "my voice assistant is sincere and genuine." Considering the Chinese language context of this investigation, all items were first translated into Chinese. Then we asked third-party experts to translate the Chinese version back to an English version. There is no significant difference between the original and the back-translated English version.

Although we observed several scales in the technology humanness literature that may be useful to measure the overall perceived humanness of conversational AI (Cho, Molina, & Wang, 2019; Westerman, Edwards, Edwards, Luo, & Spence, 2020; Westerman et al., 2019), none of them can be directly adopted to measure the two specific types of AI humanness (i.e., voice humanness and understanding humanness) posited in this study. As a result, we developed the measurement items for these two humanness constructs by following several procedural recommendations concerning scale development (Hinkin, 1995; MacKenzie, Podsakoff, & Podsakoff, 2011).

First, we conducted qualitative interviews with twenty-six users of conversational AI, each of them has over one-year of usage experience, to detect the content domains for conversational AI humanness. After given the definition and instances of conversational AI, the participants were required to answer the following questions sequentially: (1) What is the conversational AI you use most frequently? (2) What human-related features your conversational AI have? (3) Do you feel that your conversational AI is like a human in the speaking aspect? Why? (4) Do you feel that your conversational AI is like a human in the listening aspect? Why? To identify the content domains for conversational AI humanness, we categorized the responses to the second question above

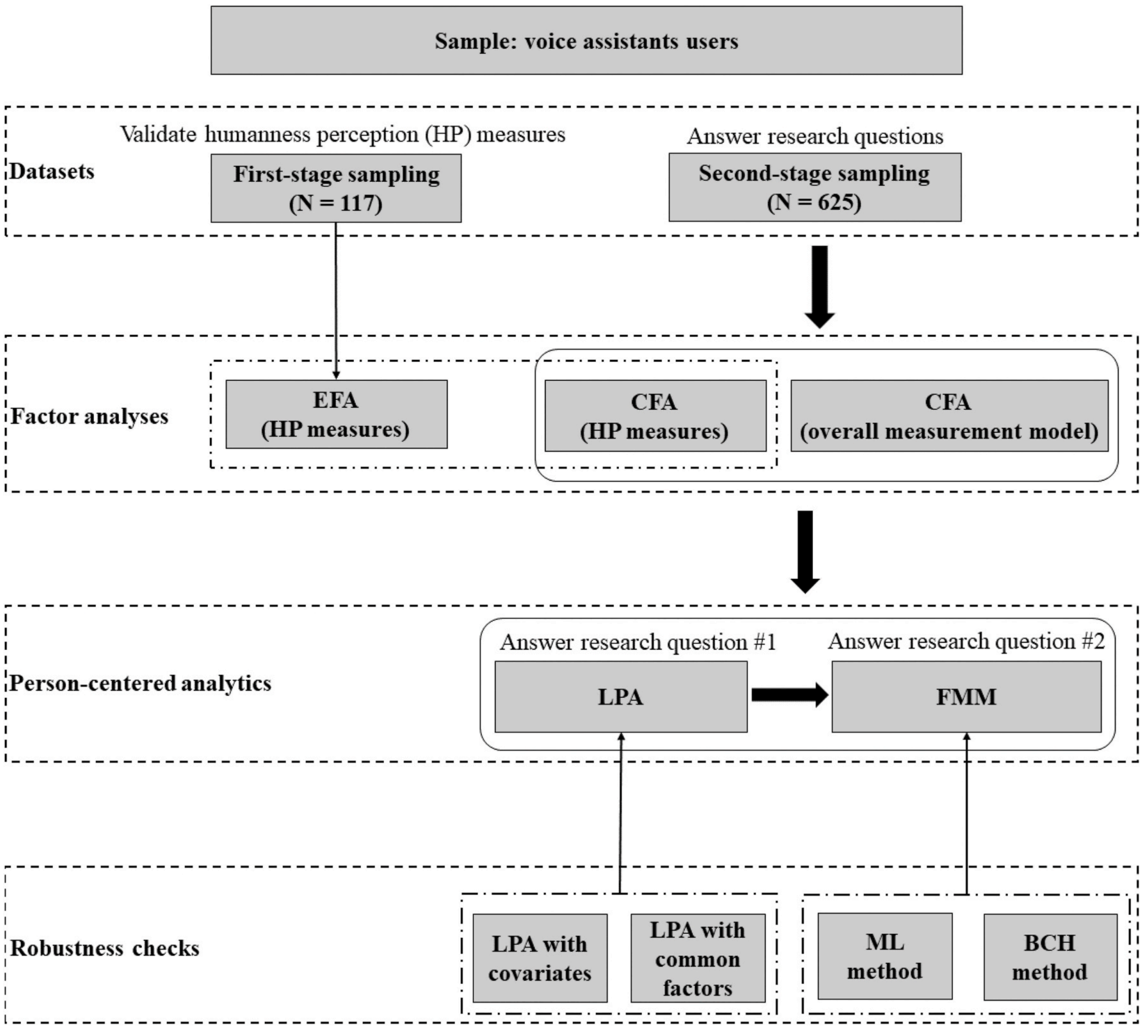

Fig. 1. Research process.

and obtained two specific humanness domains. The responses in the first domain depict how AI speaks like a human, such as "it speaks with rhythm", "having a natural voice". This domain corresponds to the voice humanness in our framework. The responses in the second domain depict how AI listens like a human, such as "good comprehension, but also misunderstand my intention sometimes". This domain corresponds to the understanding humanness in our framework. Therefore, the results of this pilot study support the proposed dualistic humanness model.

As a note, the inductive parts of this paper are the identification of AI humanness perception patterns and the trust difference among differing patterns. But, the generation of the two humanness dimensions is not an inductive process in essence, because the dual dimensionality is previously informed by the communication competence theory which posits that for an effective speech communication the interlocutors must be competent in both speaking and listening (Haas & Arnold, 1995). Here, the pilot study is used to confirm that decomposing the overall technology humanness into voice humanness and understanding humanness in the context of conversational AI is appropriate, but not to induce humanness dimension without prior information.

Second, we form the operational definitions for voice humanness perception and understanding humanness perception after the pilot study. Voice humanness perception is defined as the extent to which a user feels that a conversational AI system has a natural voice expression. Understanding humanness perception is defined as the extent to which a user feels that a conversational AI system can understand what he/she says. Because the responses to the third and fourth questions provide potential content domains for understanding humanness and voice humanness, respectively, we combine these responses, the operational definitions, and prior research pertains to technology humanness (Go & Sundar, 2019; Lankton et al., 2015; Westerman et al., 2020) to develop items for the two humanness constructs. Initially, we developed six items for each of the two humanness. Then we invited fifteen users of voice assistants to comment on the readability and clarity of these items to confirm the face validity. In this stage, no item was deleted as all participated users agreed that the meaning of each item seems to align with the measurement goal. But the wording of items was advised to make a shared adjustment by nine of the user panel. This adjustment is adding "My" before "voice assistant" in each item to ensure that respondents are indeed evaluating their own conversational AI, but not others'.

Third, after the wording adjustments, we further invited three IS researchers who specialized in user experience research to examine the content validity of the humanness measures. In this stage, one of the expert panels commented that the item "I often feel that there is a human behind my voice assistant speaking to me" captures voice but also understanding humanness and thus is unclear what humanness it intended to measure. Another external researcher pointed out the potential overlapping issue among items for understanding humanness and suggested to remove the item "My voice assistant has a human-level comprehension performance". We benefited a lot from the feedback from the expert panel and eliminated the two items above that may threaten the content validity.

Finally, we have five items for each humanness construct before collecting quantitative data to establish the construct validity. After going through cross-validation by two independent samples, the ten items for conversational AI humanness scale are retained for hypothesis testing. The details on the data collection of the two samples are described in the next section, and the results of the cross-validation are reported in the first part of the results section. Measurement items for constructs in this study are presented in Table 1. All items were measured using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree).

4.2. Data collection

The data were collected through a two-stage online survey administered on the Wenjuanxing planform. Wenjuanxing (www.wjx.cn) is the largest survey platform for market investigation and academic research in China, with more than 2.6 million registered users (Yang, Jiang, Yao, Chen, & Wei, 2018). This platform maintains panels that are representative of the Chinese population and selects participants based on the survey initiators' requirements. We recruited the company possessing the platform to distribute the link of our questionnaire to qualified members of the platform. In the service commitment statements, the company commits that they will protect the personal privacy of survey respondents, and all respondents keep anonymous during the data collection. The company also promises that only the survey initiators have the right to see the responses of participants, and they will not send our data to third parties. Overall, we choose this platform and its sampling service because of its high potential in capturing representative samples and of complying with the research ethics.

Table 1
Measurement items.

Constructs	No.	Items
Trust (competence, benevolence, integrity) (Lankton et al., 2015; McKnight et al., 2002)	TC_1	My voice assistant is competent and effective in its interactions with me.
	TC_2	My voice assistant performs all of its roles very well.
	TC_3	My voice assistant is capable and proficient.
	TC_4	In general, my voice assistant is informative.
	TB_1	I believe that my voice assistant would act in my best interest.
	TB_2	If I ask for help, my voice assistant would do its best to assist me.
	TB_3	My voice assistant is interested in my well-being.
	TI_1	My voice assistant is truthful in its dealings with me.
	TI_2	I would characterize my voice assistant as honest.
	TI_3	My voice assistant is sincere and genuine.
Voice Humanness (Go & Sundar, 2019; Schuetzler et al., 2020; Westerman et al., 2020)	VH1	My voice assistant's pronunciation is natural.
	VH2	My voice assistant has a human-like voice.
	VH3	The language expression of my voice assistant sounds like that of a machine.
	VH4	I cannot feel the distance between the voice of my voice assistant and that of a human being.
	VH5	I characterize my voice assistant's speaking aspect as human.
Understanding Humanness (Braun et al., 2019; Go & Sundar, 2019; Westerman et al., 2020)Figure A1	UH1	My voice assistant can accurately comprehend what I say.
	UH2	My voice assistant is stupid when it comes to understanding my intentions.
	UH3	The understanding ability of my voice assistant is similar to that of a human being.
	UH4	My voice assistant always misunderstands my words.
	UH5	I characterize my voice assistant's listening aspect as human-like.

In this study, the qualified members are users of conversational AI. Because voice assistants are the dominant business applications of conversational AI and the prevalent devices powered by conversational AI in users' lives (de Barcelos Silva et al., 2020), we selected the users of voice assistants as our research sample. Voice assistants have been integrated into many existing software or hardware products, such as smartphones, apps, speakers, and cars. At the beginning of the survey, we added a screening question "Have you ever used any voice assistant before?" to terminate the subjects who were not users of voice assistants. Respondents who completed and submitted the questionnaire successfully would receive a monetary reward in return. However, they can also reject the survey invitation for any reason, thus the sampling of this study, just like any other online survey research, may be subject to self-selection bias. In this sense, the sample of this study can be described as a purposeful sample.

In the first stage, we collected a sample to examine the quality of humanness perception measures (i.e., VHP and UHP). 130 users participated in this stage, and 117 valid respondents were retained for data analysis after filtering out invalid responses. In the second stage, we collected another sample to further confirm the reliability and validity of the developed humanness perception measures, and to answer the research questions of this study. 708 users participated in this stage, and 625 valid respondents were retained for data analysis after filtering out invalid responses. Therein, invalid responses were identified using the following rules: (1) completion time beyond 3 standard deviations of average level; (2) unusual responses to reversed items; (3) the same responses to all items (Meade & Craig, 2012). In the collection process, we also collected demographic data such as gender, age, education, income, in-use voice assistants, and length of usage, besides the main constructs in this study. Demographic information of the first-stage sample is presented in Appendix A, and demographic information of

Table 2
Demographic statistics of survey respondents (n = 625).

Variables	Levels	Frequency	Percentage (%)
Gender	Male	333	53.3
	Female	292	46.7
Age	<18	4	0.7
	18–25	194	31.0
	26–35	313	50.1
	36–50	105	16.8
	>50	9	1.4
Education	High school or below	58	9.3
	College's degree	132	21.1
	Bachelor degree	347	55.5
	Master's degree or higher	88	14.1
Monthly income (RMB)	<3000	119	19.0
	≥3000 but <5000	104	16.6
	≥5000 but <10000	273	43.7
	≥10000 but ≤30000	122	19.5
	>30000	7	1.1
Voice assistants	Smartphone assistants	242	38.7
	App assistants	57	9.1
	Smart speakers	230	36.8
	In-car assistants	73	11.7
	Others	23	3.7
Length of usage	<3 months	27	4.3
	≥3 but <6 months	58	9.3
	≥6 but <12 months	166	26.6
	≥1 but ≤2 years	222	35.5
	>2 years	152	24.3

Notes: A college degree program is shorter than a bachelor degree program, and the social recognition of college degrees is lower than that of bachelor degrees.

the second-stage sample is shown in Table 2.

In terms of the sample representativeness issue, we examine whether the second-stage sample can represent the user population of conversational AI in China because this sample is used to answer our research questions. By comparing the demographic information of the survey sample with the user profile of conversation AI provided by a leading market research company (www.iresearch.com.cn), we find no significant difference between the survey sample and user population of conversational AI in terms of gender ratio ($\chi^2 = 1.483$, $df = 1$, $p = 0.223$), age structure ($\chi^2 = 3.392$, $df = 4$, $p = 0.494$), education level ($\chi^2 = 5.339$, $df = 3$, $p = 0.149$), and monthly income ($\chi^2 = 4.489$, $df = 4$, $p = 0.344$). As a result, there is no salient sampling bias in this study, and the survey sample has considerable representativeness in examining humanness perception and the trust issue among the users of conversational AI.

4.3. Analytic Strategy

To answer research question 1, we use latent profile analysis (LPA) to identify the potentially distinctive perception patterns in users of conversational AI, based on their responses to measures of humanness perception of conversational AI. LPA is a popular person-centered statistical method for detecting unobserved population heterogeneity (Peugh & Fan, 2013), which has been used in management and psychology studies (Gabriel, Daniels, Diefendorff, & Greguras, 2015; Specht, Luhmann, & Geiser, 2014). Although traditional cluster analysis shares the same objective as LPA, LPA has several advantages over traditional cluster analysis. First, LPA is a model-based method that allows flexible model specification, such as including covariates or relaxing conditional independence assumption. Second, LPA is not sensitive to the measurement scale of variables, thus variables with different scale types (e.g., continuous, ordinal, and categorical) can be simultaneously included in the same model. Third, LPA is a probability-based method (not a distance-based method) that considers classification errors when classifying individuals into specific groups. Finally, LPA provides a set of formal statistical indices to determine the appropriate number of subgroups (Morin, Morizot, Boudrias, & Madore, 2011).

LPA is an inductive approach in nature, thereby several potential profile models (1–6 in this study) were estimated. We began by specifying one profile (i.e., no heterogeneity) and then successively increased the number of latent profiles until the model fit can no longer be improved by adding another profile (Nylund, Asparouhov, & Muthén, 2007). Consistent with LPA studies (Gabriel et al., 2015; Morin et al., 2011; Specht et al., 2014), seven fit indices were used to determine the number of profiles: log-likelihood (LL), Akaike information criterion (AIC), Bayesian information criterion (BIC), sample-size-adjusted BIC (SSA-BIC), Lo-Mendell-Rubin likelihood ratio test (LMR), bootstrap likelihood ratio test (BLRT), and entropy. The ideal profile model contains smaller AIC, BIC, and SSA-BIC statistics compared with other models, and an entropy value that is larger than that of other models and greater than 0.70 for classification accuracy, and significant LMR and BLRT statistics. Besides these statistical criteria above, we also consider model parsimony, model interpretability, and profile meaningfulness when determining the optimal number of profiles (Nylund et al., 2007).

After identifying the hidden perception patterns in our data using LPA, we further deploy finite mixture modeling (FMM) to analyze the relationship between perception patterns and trust in conversational AI. FMM is a modeling technique that assumes a fixed and finite number of heterogeneous subgroups in an observed data, and that allows researchers to examine predictors or outcomes of the discovered profile membership variable (i.e., perceptual patterns in this study). In the current study, we treated the three components of trust in conversational AI as outcomes of humanness perception patterns, included the components into a mixture model once we obtained an acceptable LPA model.

Particularly, we followed the automatic three-step procedure to perform the FMM analysis (Asparouhov & Muthen, 2014). First, we identified an LPA model with the optimal number of profiles based upon the criteria previously mentioned. Second, we obtained the most-likely profile membership of individuals based on the posterior probabilities estimated in the first step. Finally, we included outcome variables in the final LPA model. The most-likely class membership and classification error rate are considered when comparing outcomes between profiles, which is different from the traditional cluster analysis (Bakk & Vermunt, 2016). To model the trust components (i.e., competence, benevolence, and integrity) as outcome variables in FMM, we used the LTB method (Lanza, Tan, & Bray, 2013), which provides the comparisons among profiles on each outcome variable modeled (i.e., testing whether each profile significantly differs from each other on each outcome variable separately). Since the modeling methods for incorporating outcome variables into the mixture model have not reached a consensus in the current methodological literature, we also used two other statistical methods (i.e., ML and BCH) to test the robustness of results from the LTB method.

5. Results

5.1. Reliability and validity of measures

In this section, we use two datasets to establish the factor structure of humanness perception at first. Specifically, we use the first-stage sample as a calibration one to explore the factor structure of the construct with exploratory factor analysis (EFA). The KMO measure of sampling adequacy was equal to 0.892, and Bartlett's test of sphericity indicated significance ($p < 0.001$). Thus, the data are suitable for EFA (Fabrigar, Wegener, MacCallum, & Strahan, 1999). The results of EFA are shown in Table A2 of Appendix B. Two factors were extracted from the data, and together explained 66.850% of the total variance, beyond the cut-off 50% (Henson & Roberts, 2006). The items of factor_1 describe the humanness evaluation on the speaking aspect of voice assistants, thus factor_1 corresponds to "voice humanness". The items of factor_2 describe the humanness evaluation related to the language understanding aspect of voice assistants, thereby factor_2 corresponds to "understanding humanness". All factor loadings of items on their intended factor are higher than 0.50, with no cross-loadings were observed. Overall, the two-factor solution indicates that the two types of humanness perception can be differentiated from each other. The correspondence between items and factors suggests the acceptable quality of those items.

We then treat the second-stage sample as a validation sample to verify the factor structure and evaluate the quality of this scale with confirmative factor analysis (CFA). CFA was performed using robust maximum likelihood estimator. The model fit indices were listed as follows: $\chi^2 = 158.696$, $df = 34$, $\chi^2/df = 4.668$, CFI $= 0.955$, TLI $= 0.940$, RMSEA $= 0.077$, suggesting the measurement model of humanness perception has acceptable model fit (Evermann & Tate, 2011). All factor loadings are higher than 0.500, ranging from 0.656 to 0.819. The correlation coefficient between voice humanness and understanding humanness is 0.521, indicating that the two aspects of humanness perception are correlated at an intermediate level (see Figure A1 in Appendix B for detail).

Finally, we run a CFA including all variables used in this study to examine the overall measurement model of this study and evaluate the reliability and validity of measures within the CFA model. A favorable model fit goodness was observed: $\chi^2 = 413.390$, $df = 160$, $\chi^2/df = 2.584$, CFI $= 0.952$, TLI $= 0.943$, RMSEA $= 0.050$. Figure A2 in Appendix C shows the CFA model with standardized estimates. The reliability of measures is determined by Cronbach's alpha and composite reliability (CR). As shown in Table 3, all Cronbach's alpha values of the five variables exceed 0.70, and all CR values are higher than 0.70 as well. Thereby, the reliability of measures is favored. The validity of measures

Table 3
Reliability and validity of measures.

Variables	Cronbach's alpha	AVE	CR
1. Voice Humanness	0.878	0.599	0.882
2. Understanding Humanness	0.865	0.570	0.868
3. Competence	0.816	0.534	0.820
4. Benevolence	0.793	0.567	0.796
5. Integrity	0.799	0.574	0.801

is assessed from convergent validity and discriminant validity (Fornell & Larcker, 1981). All the average variance extracted (AVE) values of the studied variables surpass 0.50, suggesting acceptable convergent validity. Furthermore, for each variable, the square root of the AVE value is higher than its correlation coefficients with other variables. Thus, the discriminant validity of measures is supported. We also compare several potential competitive models to evaluate the distinction among variables in our measurement model (see Table 4). The results show that the five-factor model is the best-fitted one among those competitive models, which signifies that these variables are distinguished from each other, thus further supports the discriminant validity. Overall, these results suggest that the reliability and validity of measures used in the present study are acceptable.

5.2. Heterogeneous perception patterns detected by latent profile analysis

Table 5 shows the means, standard deviations, and correlations of variables in this study. The mean difference combined with the moderate magnitude of correlation between voice humanness and understanding humanness provides the evidence that the two humanness perceptions are not the same concept in users' minds. The intermediate correlations among competence, benevolence, and integrity indicate that distinguishing these three components is also necessary, although they all belong to the framework of trust.

Table 6 shows the model fit statistics for possible latent profile models (one-profile model to six-profile model). We choose the three-profile model for the following reasons: (1) although absolute LL, AIC, BIC, and SSA-BIC values decrease with the increase of profile number, they do not sharply decline any longer when more than three profiles are specified; (2) the LMR test becomes not significant from 4-profile solution, which means that model fit goodness can no longer be improved significantly when setting more than three profiles; (3) the entropy (classification accuracy) of the three-profile model (0.932) is only lower than that of the two-profile model which is however not favored by all other indices. The BLRT test keeps significant across all profile models, thus it cannot provide useful information during model selection in this case. We stopped at the six-profile model because continuing adds the number of profiles would violate the principle of both model parsimony and interpretability (Nylund et al., 2007). Consequently, we choose the three-profile model as the optimal model. In other words, three types of perception patterns were uncovered here, which we articulate in the following.

Table 7 provides the estimated means and standard errors of the LPA

Table 4
Model fit indices of competitive models.

Model	χ^2/df	CFI	TLI	RMSEA
1 factor (VH + UH + competence + benevolence + integrity)	16.322	0.511	0.453	0.157
2 factors (VH + UH, competence + benevolence + integrity)	11.366	0.671	0.630	0.129
3 factors (VH + UH, competence, benevolence + integrity)	9.499	0.733	0.697	0.117
4 factors (VH + UH, competence, benevolence, integrity)	7.826	0.790	0.756	0.105
5 factors (VH, UH, competence, benevolence, integrity)	2.584	0.952	0.943	0.050

indicators (i.e. items of humanness perception) for each profile, and Fig. 2 displays the latent response patterns of each profile.

(1) Profile 1 is characterized by high scores on indicators of both voice humanness and understanding humanness (approximately equal to 4), thus we label this profile as "para-human perception" to describe that users in this profile are prone to have an overall human-like conversation experience when interacting with their conversational AI. 45.3% of the samples are in this profile.
(2) Profile 2 is featured by low scores on indicators of both voice humanness and understanding humanness (approximately equal to 2.5), thus we label this profile as "para-machine perception" to show that users who belong to this profile tend to have an overall machine-like impression of their conversational AI in the interaction process. 29.0% of the samples are in this profile.
(3) Different from the two profiles above, Profile 3 exhibits high scores on indicators of voice humanness but low scores on indicators of understanding humanness. The Wald test demonstrates that for this profile, voice humanness (M = 3.901) is significantly higher than understanding humanness (M = 2.201), $p < 0.001$. Thereby, we label this profile as "asymmetric perception" to outline the asymmetry between the two humanness perceptions for users in the profile, and 25.7% of the samples are in this profile. These results indicate the existence of heterogeneous perception patterns for conversational AI's humanness, which answers Research Question 1. Next, we further explore the implications of the revealed perception patterns for trust in conversation AI, to answer Research Question 2.

5.3. Linking perception heterogeneity with user trust by finite mixture modeling

We run a finite mixture model with trust facets as the outcomes to examine whether distinct perception patterns have different implications for user trust. Table 8 presents the results of the finite mixture model. In general, all three facets can be differentiated by perception patterns, supported by the overall χ^2 tests with $p < 0.001$.

Specifically, for competence, users in the profile of para-human perception exhibit a higher level of competence-based trust than users in other profiles, suggesting a positive implication of dual human-like sense for user trust in terms of the ability of conversational AI. However, the competence difference between profiles of para-machine perception and asymmetric perception is not significant, indicating that voice humanness of conversational AI may no longer favor users' competence-related trust once understanding humanness is limited at a low level, because the two profiles share low understanding humanness but separate at voice humanness.

As to benevolence, the results also reveal a positive implication of dual human-like sense for user trust in conversational AI since users with para-human perception show a higher level of benevolence-based trust than users with other perception patterns. Surprisingly, unlike competence, users with asymmetric perception demonstrate a lower benevolence-based trust that users with para-machine perception. In other words, the improvement in voice humanness may, instead of not enhancing, disintegrate user trust related to conversational AI's benevolence when understanding humanness is confined at a low level. On the other hand, switching the reference from para-machine pattern to para-human pattern, the results show that the reduction in understanding humanness could undermine users' benevolence-based trust despite high voice humanness is already possessed. These results signify that the asymmetry itself may impair user trust in the benevolence aspect of conversational AI.

In terms of integrity, we obtain similar results to that of benevolence. The users with asymmetric perception exhibit the lowest integrity-related trust among the three profiles. Indeed, both benevolence and integrity are associated with the morality of humans (Sonpar,

Table 5
Descriptive statistics (n = 625).

Variables	M	SD	1	2	3	4	5
1. Voice Humanness	3.470	0.919	**0.774**				
2. Understanding Humanness	2.935	0.884	0.525	**0.755**			
3. Competence	3.621	0.779	0.373	0.393	**0.731**		
4. Benevolence	3.340	0.826	0.216	0.387	0.476	**0.753**	
5. Integrity	3.593	0.774	0.206	0.438	0.571	0.544	**0.757**

Notes: The bolded diagonal elements are the square roots of the average variance extracted (AVE) values.

Table 6
Fit statistics for latent profile models (n = 625).

No. of profiles	LL	FP	AIC	BIC	SSA-BIC	LMR (*p*)	BLRT (*p*)	Entropy
1	−9481.359	20	19002.718	19091.473	19027.976	–	–	–
2	−8479.298	31	17020.597	17158.167	17059.746	0.000	0.000	0.937
3	−7964.827	42	16013.654	16200.040	16066.696	0.000	0.000	0.932
4	−7855.228	53	15816.456	16051.657	15883.389	0.271	0.000	0.902
5	−7794.293	64	15716.586	16000.602	15797.411	0.500	0.000	0.846
6	−7709.841	75	15569.683	15902.514	15664.400	0.304	0.000	0.851

Note: LL = log-likelihood; FP = the number of free parameters; AIC = Akaike information criteria; BIC = Bayesian information criteria; SSA-BIC = sample-size adjusted BIC; LMR = Lo, Mendell, and Rubin (2001) test; BLRT = bootstrapped log-likelihood ratio test.

Table 7
Parameter estimates for the distinct profiles.

Indicators/ Profiles	Profile 1 Para-human perception	Profile 2 Para-machine perception	Profile 3 Asymmetric perception
Voice humanness			
VH1	3.982 (0.033)	2.400 (0.080)	4.100 (0.053)
VH2	3.963 (0.053)	2.393 (0.078)	3.924 (0.079)
VH3	3.994 (0.053)	2.161 (0.049)	3.808 (0.091)
VH4	4.027 (0.056)	1.950 (0.058)	3.796 (0.098)
VH5	4.099 (0.050)	2.290 (0.077)	3.906 (0.087)
Understanding humanness			
UH1	3.855 (0.033)	2.442 (0.069)	2.221 (0.076)
UH2	3.903 (0.039)	2.395 (0.076)	2.097 (0.082)
UH3	3.629 (0.065)	2.265 (0.068)	2.358 (0.078)
UH4	3.630 (0.070)	2.038 (0.057)	2.079 (0.060)
UH5	3.775 (0.050)	2.360 (0.076)	2.250 (0.089)
% of sample	45.3%	29.0%	25.7%

Notes: The values in parentheses are the standard errors of estimated means.

Handelman, & Dastmalchian, 2009). Therefore, the results indicate a negative implication of asymmetric humanness perception for moral-related trust in conversational AI. In sum, by finite mixture modeling, we uncover complex (positive and negative) relationships between humanness perception heterogeneity and varied trust facets, in response to Research Question 2.

5.4. Robustness checks

We performed several additional analyses to check the robustness of results answering Research Question 1 and Research Question 2, respectively. In response to Research Question 1, three distinct profiles were uncovered in which a profile with asymmetric perception emerged. To verify the stability of the profile structure, we conducted two additional analyses: (1) controlling for covariate effect by incorporating covariates into the LPA model; (2) relaxing the assumption of conditional independence by adding two common factors to the LPA model.

First, if the number and configuration of profiles are robust, the inclusion of covariates can affect, if any, only class probabilities (Marsh, Lüdtke, Trautwein, & Morin, 2009). We thus included the four demographic and two conversational AI usage variables (see Table 2) in the LPA model to examine whether the profile structure we discovered above is dependent on the existence of covariates. The results of the LPA model with covariates are shown in Table A3 in Appendix D. The configuration of profiles is displayed in Figure A3 in Appendix D. The inclusion of covariates produces similar results with the original LPA model. Although the LMR test becomes significant in the six-profile model, other fit indicators such as entropy and information criteria still favor the three-profile model, as well as when model parsimony and interpretability are considered. Moreover, the specific response patterns of profiles and their sample distribution are also similar before and after including covariates (see Figure A3). Therefore, the results as to profile structure (i.e., number and feature) are not conditional on the existence of covariates.

Second, the LPA assumes that the correlations among indicators can be sufficiently explained by a latent categorical variable (Peugh & Fan, 2013). That is to say, the indicators may be independent of each other once a latent categorical variable is specified. However, this assumption is often too strict with real data, especially when the indicators used in the LPA model can be explained by latent continuous variables in theory as well (Morin et al., 2011). In our case, the indicators used to define latent profiles are the items of voice humanness and understanding humanness. Thus, theoretically, it is plausible to add two common factors to the LPA model, with one explaining the correlations among items of voice humanness and the other accounting for the correlations among items of understanding humanness. The fit indices of the LPA model with common factors are presented in Table A4 in Appendix D. The significant LMR test of the four-profile model seems to support this model against the three-profile one. However, by checking the generated profile configuration (see Figure A5 in Appendix D), we observed a profile (Profile 3) that cannot be explained in theory, especially the shake within the items of voice humanness, and that only comprising 9.5% of our sample. By contrasting the profile configurations between these two models (Fig. 2 and Figure A5), we find that the abnormal profile is almost a sub-profile of the para-human perception profile in the three-profile model. Consequently, the three-profile model was favored as the best one. Moreover, Figure A4 shows a similar profile configuration with Fig. 2. In sum, the relaxation of conditional independence has not altered the number and nature of profiles, thus justifies the robustness of the profile structure of humanness perception again.

Finally, to answer Research Question 2, we have deployed the LTB method to perform the finite mixture model before. However, the issue of how best to model distal outcomes in mixture models is an ongoing discussion in the methodological literature. Indeed, the ML method (Nylund-Gibson, Grimm, Quirk, & Furlong, 2014) and the BCH method (Asparouhov & Muthen, 2014) are also proposed as effective alternatives to handle distal outcomes in mixture models. As, till now, the

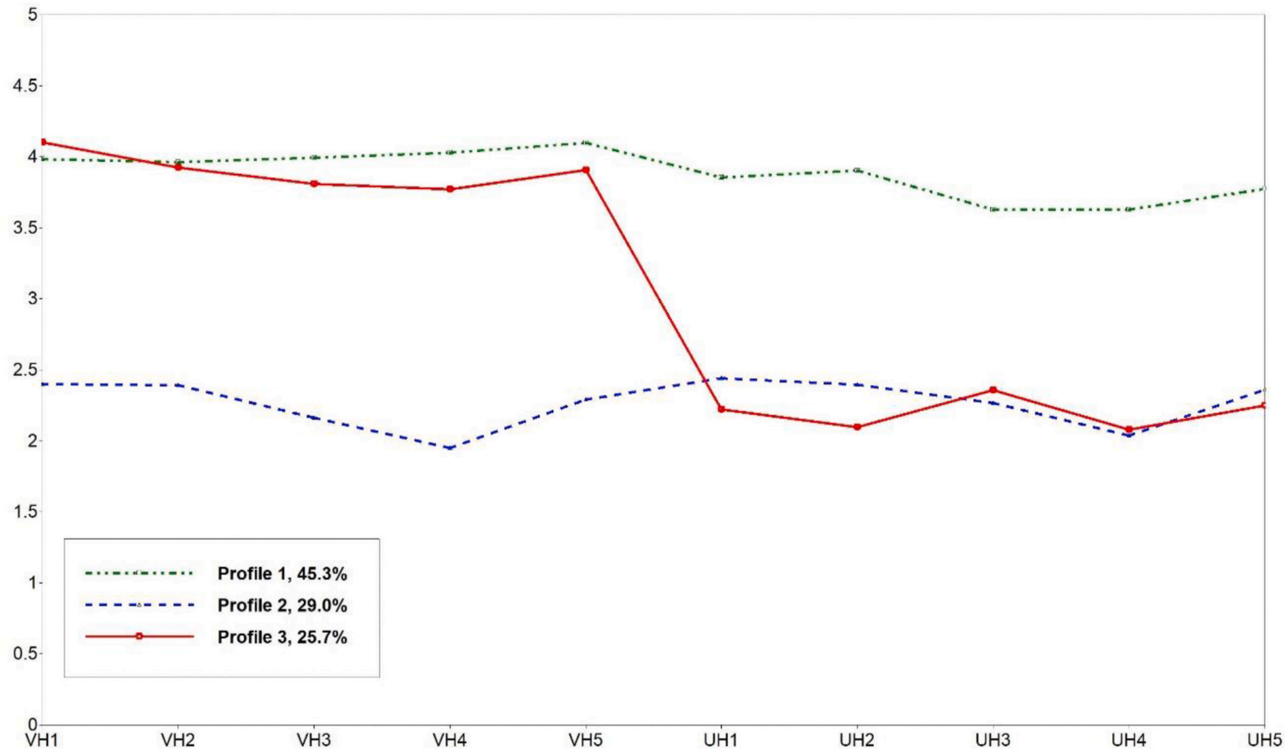

Fig. 2. Latent Profiles for Heterogeneous Humanness Perception. Notes: VH1-VH5 and UH1-UH5 are the items of voice humanness and understanding humanness, respectively.

Table 8
Results for the finite mixture model.

Trust facets	1. Para-human perception	2. Para-machine perception	3. Asymmetric perception	Overall χ^2	Differences among profiles
Competence	3.930 (0.036)	3.338 (0.061)	3.404 (0.063)	97.709***	**1 > 2 = 3**
Benevolence	3.643 (0.045)	3.231 (0.059)	2.951 (0.061)	88.902***	**1 > 2 > 3**
Integrity	3.949 (0.036)	3.481 (0.054)	3.135 (0.062)	148.428***	**1 > 2 > 3**

Notes: ">" denotes differences between two profiles as to trust facets are significant at $p < 0.05$, whereas " = " denotes the differences above are not significant at $p <$ 0.05; ***$p < 0.001$.

superiority among these modeling approaches has not reached a consensus, we thereby also utilize the two methods above to model the trust facets in our finite mixture model. The results are summarized in Table A5 and Table A6. The findings discovered by the ML method and the BCH method are the same as those unearthed by the LTB method. Therefore, our findings of the implications of humanness perception patterns for trust in conversational AI are robust to modeling methods.

In summary, the answers to the research questions are not sensitive to the existence of covariates. In addition, they do not rely on a model based on a stringent assumption, and they are not dependent on a specific statistical method as well.

6. Discussion

6.1. Summary of key findings

The recent advances in natural language processing nudge the advent of conversation AI that can converse with users in a human-like way and thus transforms the interactional way between users and systems. At the same time, the absence of trust in conversational AI has been documented by many industrial reports and market investigations (Martin, 2019; Olson, 2019; Schwartz, 2020). This study aims to contribute new knowledge about the trust issue of conversational AI from the humanness perception angle by using a person-centered approach. Particularly, we examine the existence of heterogeneous perception patterns based on the proposed dualistic humanness for conversation AI, and how the heterogeneity of humanness perceptions relates to trust in conversational AI. The findings of this paper are summarized as a theoretical model in Fig. 3. As this paper is a person-centered study, the findings may not be informed only by viewing the theoretical model. Thus, in the following, we discuss the findings and elaborate on the theoretical model in detail.

First, drawing on communicative competence theory, this study

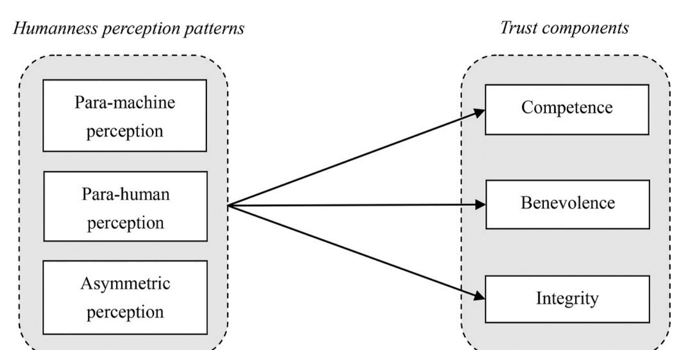

Fig. 3. The relationships between humanness perception patterns and trust components.

decomposes the humanness of conversational AI into voice humanness and understanding humanness and proposes a dualistic humanness model for conversational AI. Based on this model, we discover three distinct humanness perception patterns, which are displayed on the left side of Fig. 3. The discovery of para-machine and para-human pattern is not uncommon as the literature on technology humanness revealed that technologies can be perceived differently in the dimension of humanness and the degree of humanness perception is mainly dependent on the humanness cues exhibited by technologies (Califf, Brooks, & Longstreet, 2020; Lankton et al., 2015). Thereby, some users can perceive overall low humanness (i.e., para-machine pattern) while others can perceive overall high humanness (i.e., para-human pattern), in voice interaction with their conversational AI. However, we also uncover an asymmetric perception pattern with high perceived voice humanness but low understanding humanness. This pattern deepens our understanding of the humanness nature of conversational AI since it indicates that users can perceive different dimensions of the humanness of an AI system diversely. While prior studies focused on the overall humanness perception and its effect on user behavior (Go & Sundar, 2019; Lankton et al., 2015; Westerman et al., 2020), this study suggests that disaggregating the humanness cues embedded in an AI system could be favorable for knowing the relative value of individual humanness factors (Nass & Moon, 2000).

Second, we find that the users' trust level towards conversational AI is associated with the specific perception patterns they have. Particularly, for users who have a para-human perception regarding conversational AI, they show the highest trust level on all the three components among the discovered perception patterns. This finding is supported by previous research documenting the positive effect of perceived humanness of technology on user trust (de Visser et al., 2016; Qiu & Benbasat, 2009; Waytz et al., 2014). Interestingly, this positive effect is also receiving challenges from recent HCI research suggesting a negative effect when artificial agents are rated as too human-like (Culley & Madhavan, 2013; Shin, Kim, & Biocca, 2019). As a note, our study is not suggesting that the more humanness users perceive, the higher trust users have in conversational AI, because this inference can only be tested by a variable-centered study quantifying the relations among variables. This study only indicates that users who have a relatively high humanness perception (i.e., approximately 4 on a 5-point rating scale) on both voice and understanding of conversational AI could show greater trust towards conversational AI than the other two perception patterns.

Third, for users who have a para-machine perception regarding conversational AI, their levels of trust in conversational AI are surprisingly not the lowest among the discovered perception patterns. Instead, users who possess an asymmetric perception pattern show the lowest morality-based trust. This finding suggests that the asymmetry between humanness perceptions on the speaking and listening of conversational AI may leads users to question the ethics of their conversational AIs. The uncertainty reduction theory provides a possible explanation for this finding: the asymmetry between these two humanness triggers a perceived uncertainty about the identification or nature of conversational AI (i.e., an object between human and machine), whereas perceived uncertainty has been identified as a strong inhibitor for user trust in many contexts (Srivastava & Chandra, 2018). However, this asymmetry may not affect users' competence-based trust regarding conversational AI as we find that users with para-machine perception have a similar level of competence-based trust with users who possess asymmetric perception. The competence of a conversational AI mainly depends on its performance on spoken language understanding (Braun et al., 2019), thus the high voice humanness perception in the asymmetric pattern cannot facilitate competence-based trust.

6.2. Theoretical Contributions

First, by shifting from a unilateral relationship perspective to an interactive relationship angle, this study proposes a dualistic model of humanness for conversational AI (i.e., voice humanness and understanding humanness), which provides a novel framework for future research to evaluate users' humanness perceptions of conversational AI, and examine the impact of human-like design of conversational AI on user experience and behavior. The unilateral relationship between users and systems is a classical assumption which posits that systems have many functional features and users utilize these features to achieve a goal (Schuetz & Venkatesh, 2020). Following this assumption, previous studies examined how users perceive a system's voice that is imbued with diverse humanness cues (Chang et al., 2018; Edwards et al., 2019; Torre et al., 2020; Xu, 2019). Indeed, a human-like voice is a new feature added to systems. But the voice of systems is used to *interact* with users in spoken language. Voice interaction requires that systems should also be able to understand what users say. More importantly, when systems understand the spoken language expressions of users, systems can use the learned preferences or habits of users to do something that is not requested by users, such as actively adjusting conversational style the next time. In other words, these intelligent systems can use users to achieve their objectives, and thus the unilateral relationship assumption is challenged by interactive AI systems (Demetis & Lee, 2018; Schuetz & Venkatesh, 2020), such as conversational AI. Out of this assumption, new research problems may arise. Accordingly, this study takes an interactive angle to view the humanness phenomenon of conversational AI and proposes a dual humanness model for conversational AI: voice humanness and understanding humanness. Voice humanness corresponds to the speaking attribute of conversational AI, whereas understanding humanness corresponds to the listening attribute of conversational AI. Just like the two sides of a coin, voice humanness and understanding humanness are different from each other, but together they enable the human-AI conversation. The humanness factors of conversational AI examined in prior works mainly fall on voice humanness. In this sense, the proposed dualistic model can inspire future research to simultaneously consider this two humanness when exploring the relationship between conversational AI's humanness and user experience.

Second, this study uncovers three distinct latent profiles of humanness perceptions for conversation AI that provide the first empirical insights into what perception patterns users have when interacting with their conversation AI. Particularly, we find two quantitatively distinct perception patterns with one reflecting an overall human-like perception and the other mirroring an overall machine-like perception. Surprisingly, we do not find a perception pattern exhibiting moderate evaluation on voice humanness and understanding humanness. The absence of this profile suggests that perceived humanness may not be a continuum, but rather be dichotomous, in users' mind. Categorical processing can explain this absence, because individuals tend to cognize the objects around their lives in a categorical manner, and thus are unlikely to endure an ambiguous category when they perceive their conversational AI (McKone, Martini, & Nakayama, 2001; Wiese & Weis, 2020). This inherent processing tendency push users away from perceiving conversational AI as a bit human-like but also a bit machine-like (i.e., moderate evaluation on both voice and understanding humanness), to perceiving conversational AI as human-like or machine-like. Considering that the literature of technology humanness suggests that technologies vary in their perceived humanness (Cho et al., 2019; Lankton et al., 2015; Westerman et al., 2019), this finding contributes to this literature stream by presenting new knowledge about the nature of the concept of humanness perception. Furthermore, this finding outlines that subjective humanness perception deserves more attention when exploring the influences of objective human-like designs in AI systems on user experience because of the potential nonlinear link between objective humanness design and subjective humanness perception.

We also find a perception pattern that qualitatively distinct from the two perception patterns mentioned above. This perception pattern is featured by high voice humanness perception but low understanding humanness perception. This finding indicates that those humanness cues embedded in speaking and listening aspects of conversational AI are not assessed at a unified scale for users, thereby differentiating the two aspects in the future research could be considered. Unexpectedly, we do not find the opposite (i.e., low voice humanness perception but high understanding humanness perception) of the perception pattern above. This finding demonstrates the advantage of the person-centered approach for this study, because if employing a variable-centered approach, the undiscovered perception pattern above would be, by default, assumed to exist in the real world. The person-centered approach (e.g., LPA) produces subgroups based on heterogeneity observed in data, whereas the variable-centered approach (e.g., interaction effect testing) could produce artificial subgroups that may or may not exist, and a created subgroup that may not exist would result in misleading findings (Howard & Hoffman, 2018).

Finally, this study reveals that users with distinct perception patterns have different levels of trust in conversational AI. On one hand, we find that users with asymmetric humanness perception show a similar level of competence-based trust with users with para-machine perception. The configurational difference between these two perceptual patterns mainly falls on voice humanness. Thus, this finding suggests that enhanced voice humanness may not be able to boost the user's trust in conversation AI's ability once users feel that conversational AI cannot understand their words. That is to say, the benefits of humanizing AI's voice for user trust documented in previous research may disappear when AI systems are poor in natural language understanding (Chang et al., 2018; Edwards et al., 2019; Torre et al., 2020; Xu, 2019), because comprehending user commands accurately is a prerequisite for conversational AI to assist users. On the other hand, we also find that users with asymmetric humanness perception demonstrate the lowest level of trust as to conversational AI's benevolence and integrity. Benevolence and integrity are considered as morality-related trust (Sonpar et al., 2009). Thus, this finding indicates a negative influence of asymmetric perception between voice humanness and understanding humanness on users' morality-related trust. As discussed above, the perceived uncertainty about the identification of conversational AI might play a critical role in inhibiting the morality-related trust of users who perceive asymmetric humanness, which represents an avenue for future research to verify this possible theoretical mechanism.

6.3. Implications for practice

This study provides several implications for AI practitioners to improve different facets of user trust. First, for competence-based trust, this study finds that there is no significant difference between the para-machine perception pattern and the asymmetric perception pattern in terms of competence-based trust in conversation AI, despite that the asymmetric perception pattern characterizes high perceived voice humanness of conversational AI. This finding suggests that the path of enhancing the voice humanness of conversation AI to make users trust in conversational AI's competence may not work when the spoken language understanding of conversational AI has a poor performance. This deficiency in spoken language understanding ability is often caused by limited financial capital or restricted technological resource possessed by specific AI companies, which is especially true for AI startups. Therefore, AI companies having limited technological or capital advantages should be cautious about the decision to humanize the voice of their conversational AI products when the spoken language understanding performance of the products is relatively poor, because the investments at this moment may not be able to gain a boosted user trust in the ability of their AI products. Rather, these AI companies should concentrate all available resources to improve their AI products' speech understanding, despite that the optimization of AI's speech understanding can be more difficult than creating a human-like voice for AI.

Second, for morality-based trust (i.e., benevolence and integrity), this study uncovers that the asymmetric perception pattern, featured by high voice humanness perception but low understanding humanness perception of conversational AI, has the lowest morality-based trust among the three heterogeneous patterns. This finding suggests that the perceived asymmetry between the two humanness aspects could undermine user trust regarding the morality of conversational AI. The ethics of conversational AI may not impact its daily usage as long as it functions well in performing various tasks such as controlling home appliances or playing music. However, this factor is very important for users to do shopping-related activities via conversational AI, because the ethics issue has been identified as a crucial factor for online transaction activities by prior e-commerce research (Cheng, Yang, Chen, & Chen, 2014). In other words, people would be unwilling to accept product recommendations offered by their conversational AI when they feel that the in-use conversational AI lacks morality, let alone shopping via it. For example, users may be concerned that conversational AI uses privacy data to recommend products that maximize the benefits of the companies behind it. Nowadays, many conversational AIs have been used in product recommendations or online shopping, such as Amazon Alexa and Alibaba Tmall Genie (Klaus & Zaichkowsky, 2020). Therefore, our finding of the association between asymmetric perception and morality-based trust can be helpful for AI companies in the way to monetize the business value of conversational AI. Particularly, this study suggests that AI companies should conduct a pilot study to collect users' feedback on the speaking and listening aspects of their AI product before launching it, and then in the next stage put more investment on the weak aspect that received a relatively poor evaluation in the pilot study to develop a conversational AI which has balanced humanness on speaking and listening.

Finally, this study does not find an ambiguous pattern characterized by moderate voice humanness perception and moderate understanding humanness perception. This highlights the need for AI developers to pay more attention to users' subjective perceptions when they strive to humanize conversational AI, because the perceived humanness regarding conversational AI may not be evaluated by users in a continuous manner. In other words, adding a new humanness cue into conversational AI or reinforcing an existing human-related feature may not necessarily result in an improved humanness perception, because humans inherently tend to use categorical processing to perceive objects around them and avoid ambiguous cognitive states (McKone et al., 2001; Srivastava & Chandra, 2018; Wiese & Weis, 2020). Therefore, AI companies should seek a trade-off between the cost of human-like designs and the benefit of humanness users perceive.

6.4. Limitations and future research Direction

We acknowledge several limitations of the present study. First, by drawing on the communicative competence theory, this study proposes a dualistic humanness model for conversational AI wherein voice humanness reflects the speaking aspect of conversational AI and understanding humanness mirrors the listening aspect of conversational AI. This theoretical perspective outlines the conversation-related skills across the communication process. However, there may be other human-related factors beyond the user-AI conversation considered in the current study, such as the perceived humanness related to the personality of conversation AI (Lee, Peng, Jin, & Yan, 2006). Future research can extend our proposed humanness model by using other insightful theoretical lenses.

Second, this paper uses a few items related to the understanding performance of conversational AI to measure users' perceived understanding humanness, because the understanding humanness is rooted in how an AI performs well in understanding users' words like a human. Thus, an AI's performance level in language understanding constitutes an important indicator for the understanding humanness. Nevertheless, this operational method may also lead to the potential face validity issue. Future research can examine this construct from different theoretical viewpoints and refine its measures in the current study to accommodate the rapid development of conversational AI technology.

Finally, this study takes a person-centered approach to explore users' humanness perception patterns regarding conversational AI and its association with user trust. Although this approach can address the latent heterogeneity issue within users by detecting unobserved subgroups and relating the membership variable to other variables of interest, it is in nature an exploratory methodological angle. Thereby, the findings of the present study need to be validated by future confirmatory studies using deductive-oriented methods and samples from other countries or cultures. Our findings can also be complemented by future research that examines trust-related behaviors such as shopping via conversational AI (Rhee & Choi, 2020).

7. Conclusion

Nowadays, conversational AI can use natural language to communicate with users. This upgrade in interactive modality can bring about user trust because it indicates that AI is becoming more powerful. However, the absence of user trust has become a major concern for AI managers. This phenomenon seems interesting, but little is known about how the lack of trust occurs. This study enriches our knowledge on this issue in two ways. First, we focus on the unique features of conversational AI (i.e., speaking and listening) that differentiate conversational AI from other forms of AI systems. Second, we shift the methodological mindset from a variable-centered approach to a person-centered approach. As a result, we examine the existence of heterogeneous humanness perception patterns regarding the speaking and listening of conversation AI, and the relationship between humanness perception heterogeneity and trust in conversational AI.

We discover three patterns in conversation AI users: para-human, para-machine, and asymmetric perception, which deepen the understanding of how users distinctively perceive the humanness manifested by the speaking and listening of conversational AI. We also find that voice humanization cannot facilitate competence-related trust when AI devices' language understanding is perceived as poor. Surprisingly, users with asymmetric perception show the lowest level of morality-related trust among the discovered patterns. This finding indicates that the speaking and listening attributes of conversation AI are assessed at separated scales, but also implies that the asymmetry between voice humanness and understanding humanness can impair morality-related trust. This study suggests that AI developers should be cautious to humanize AI devices' voice and be more attentive to the potential imbalance between the speaking and listening function of AI devices.

Funding

This work was supported by the grants from National Natural Science Foundation of China (Project No. 71810107003) and National Social Science Fund of China (Project No. 18ZDA109). This work was also supported by Business Intelligence Center of EMLYON and Modern Information Management Research Center at HUST.

Credit author statement

Peng Hu: Conceptualization, Formal analysis, Writing – original draft, Yaobin Lu: Data curation, Resources, Writing – review & editing, Supervision, Yeming (Yale) GONG: Methodology, Writing – review & editing, Validation.

Appendix A. Demographic Information of the First-stage Sample

Table A1
Demographic Statistics of Survey Respondents (n = 117).

Variables	Levels	Frequency	Percentage (%)
Gender	Male	65	55.6
	Female	52	44.4
Age	<18	1	0.9
	18–25	38	32.5
	26–35	65	55.6
	36–50	13	11.1
	>50	0	0
Education	High school or below	3	2.6
	College's degree	13	11.1
	Bachelor's degree	87	74.4
	Master's degree or higher	14	12.0
Monthly income (RMB)	<3000	19	16.2
	3000–5000	17	14.5

(continued on next page)

Table A1 (*continued*)

Variables	Levels	Frequency	Percentage (%)
	5000–10000	53	45.3
	10000-30000	26	22.2
	>30000	2	1.7
Voice assistants	Smartphone assistants	47	40.2
	App assistants	7	6.0
	Smart speakers	52	44.4
	In-car assistants	8	6.8
	Others	3	2.6
Length of usage	<3 months	3	2.6
	3–6 months	10	8.5
	6–12 months	37	31.6
	1–2 years	33	28.2
	>2 years	34	29.1

Appendix B. Factor Analysis Results of Humanness Perception Measures

Table A2
EFA Results of Dual Humanness Measures (n = 117).

Items	Factor_1	Factor_2
• My voice assistant's pronunciation is natural.	0.865	0.116
• My voice assistant has a human-like voice.	0.787	0.155
• The language expression of my voice assistant sounds like that of a machine.	0.775	0.228
• I cannot feel the distance between the voice of my voice assistant and that of a human being.	0.788	0.271
• I characterize my voice assistant's speaking aspect as human.	0.776	0.227
• My voice assistant can accurately comprehend what I say.	0.174	0.831
• My voice assistant is stupid when it comes to understanding my intentions.	0.122	0.841
• The understanding ability of my voice assistant is similar to that of a human being.	0.227	0.694
• My voice assistant always misunderstands my words.	0.282	0.771
• I characterize my voice assistant's listening aspect as human-like.	0.173	0.775
Percentage of variance explained	33.966%	32.884%
KMO value	0.892	
Bartlett's test of sphericity	$p < 0.001$	

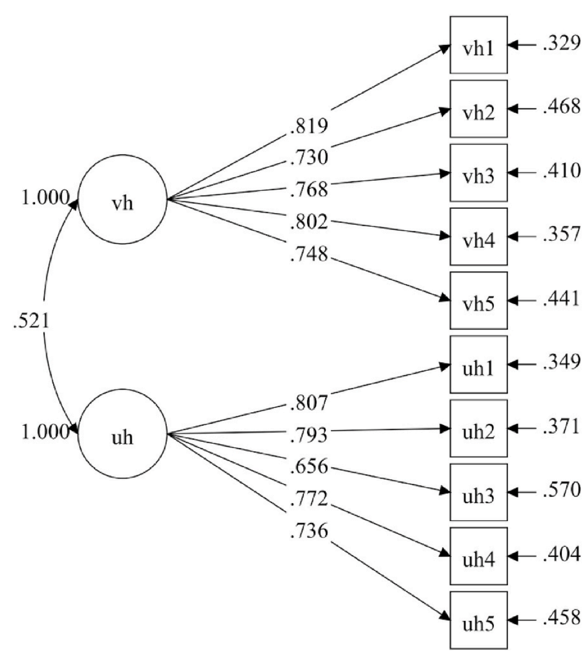

Fig. A1. Standardized Solutions of CFA on Humanness Measures (n = 625). Notes: "vh" denotes voice humanness, "uh" denotes understanding humanness.

Appendix C. Confirmative Factor Analysis for Measurement Model

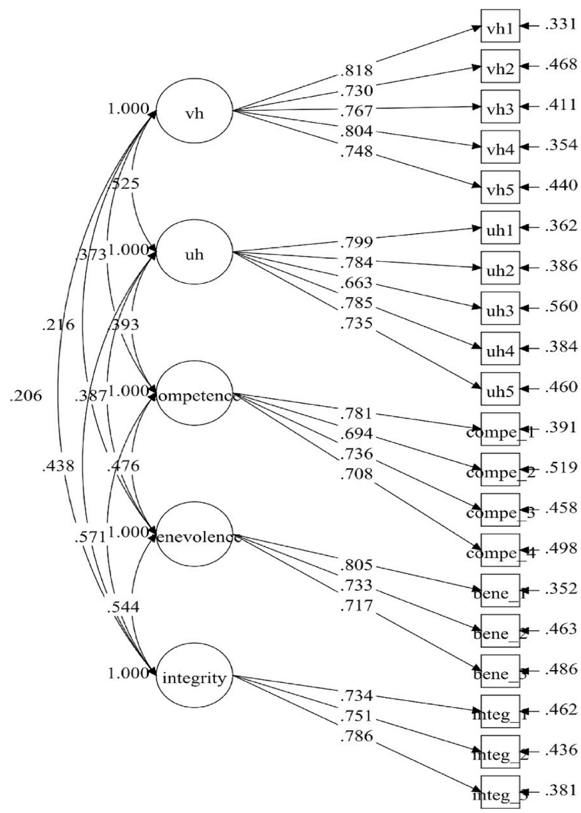

Fig. A2. CFA Model with Standardized Estimates. Notes: "vh" denotes voice humanness; "uh" denotes understanding humanness.

Appendix D. Robustness Checks for Profile Structure

Table A3
Fit Statistics for Latent Profile Models with Covariates.

# of profiles	LL	FP	AIC	BIC	SSA-BIC	LMR (p)	BLRT (p)	Entropy
1	−14205.846	31	28473.693	28611.263	28512.843	–	–	–
2	−8461.184	37	16996.367	17160.564	17043.094	0.000	0.000	0.940
3	−7941.912	54	15991.823	16231.462	16060.019	0.000	0.000	0.934
4	−7828.961	71	15799.923	16115.003	15889.588	0.134	0.000	0.905
5	−7748.999	88	15673.997	16064.520	15785.132	0.110	0.000	0.916
6	−7663.587	105	15537.174	16003.138	15669.778	0.026	0.000	0.861

Note: LL = log likelihood; FP = free parameters; AIC = Akaike information criteria; BIC = Bayesian information criteria; SSA-BIC = sample-size adjusted BIC; LMR = Lo, Mendell, and Rubin (2001) test; BLRT = bootstrapped log-likelihood ratio test.

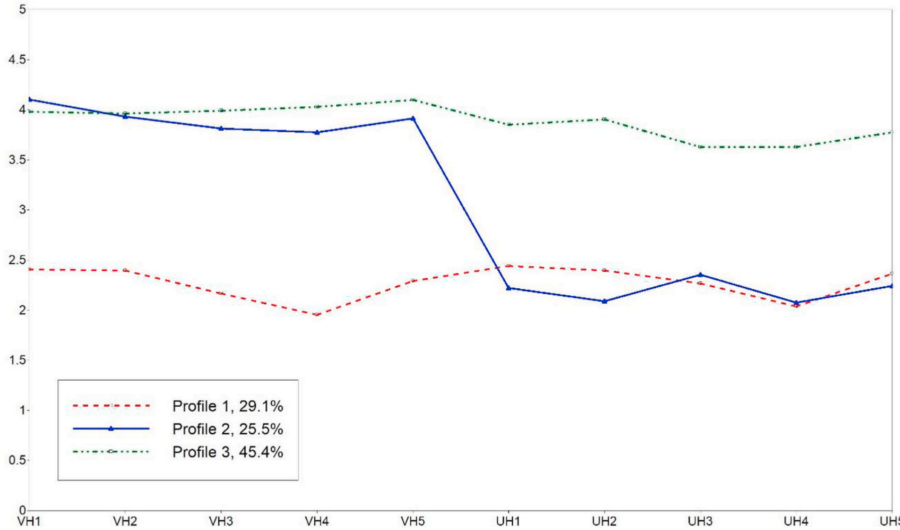

Fig. A3. Latent Profiles Disclosed by LPA with Covariates.

Table A4
Fit Statistics for Latent Profile Models with Common Factors.

# of profiles	LL	FP	AIC	BIC	SSA-BIC	LMR (*p*)	BLRT (*p*)	Entropy
1	−7948.904	31	15959.809	16097.379	15998.958	–	–	–
2	−7827.478	42	15738.957	15925.342	15791.998	0.000	0.000	0.885
3	−7744.733	53	15595.466	15830.667	15662.399	0.024	0.000	0.874
4	−7684.929	64	15497.858	15781.874	15578.683	0.021	0.000	0.842
5	−7629.713	75	15409.426	15742.257	15504.142	0.174	0.000	0.884
6	−7592.594	86	15357.188	15738.834	15465.796	0.843	0.000	0.857

Note: LL = log likelihood; FP = free parameters; AIC = Akaike information criteria; BIC = Bayesian information criteria; SSA-BIC = sample-size adjusted BIC; LMR = Lo, Mendell, and Rubin (2001) test; BLRT = bootstrapped log-likelihood ratio test.

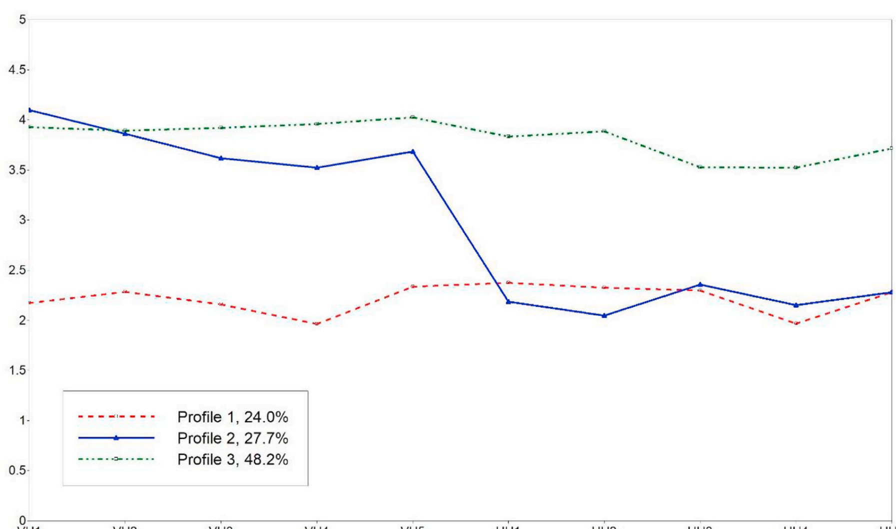

Fig. A4. Three-profile Model of LPA with Common Factors.

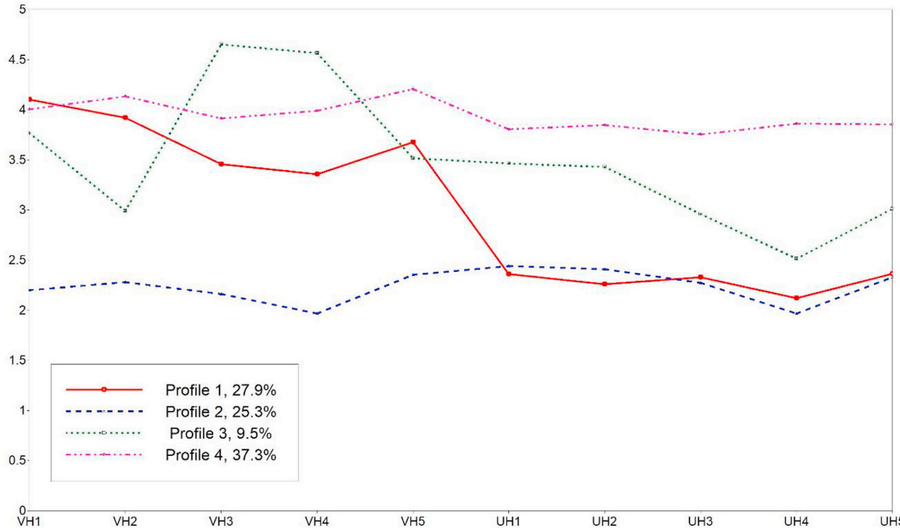

Fig. A5. Four-profile Model of LPA with Common Factors.

Appendix E. Robustness Checks for Trust Implications of Perception Pattern

Table A5
Results for the Finite Mixture Model Using the ML Method.

Trust facets	1. Para-human perception	2. Para-machine perception	3. Asymmetric perception	Overall χ^2	Differences among profiles
Competence	3.942 (0.033)	3.340 (0.062)	3.373 (0.075)	103.526***	**1 > 2 = 3**
Benevolence	3.632 (0.049)	3.244 (0.056)	2.933 (0.069)	68.750***	**1 > 2 > 3**
Integrity	3.961 (0.042)	3.477 (0.059)	3.092 (0.067)	122.375***	**1 > 2 > 3**

Notes: ">" denotes differences between two profiles as to trust facets are significant at $p < 0.05$, whereas " = " denotes the differences above are not significant at $p < 0.05$; ***$p < 0.001$.

Table A6
Results for the Finite Mixture Model Using the BCH Method.

Trust facets	1. Para-human perception	2. Para-machine perception	3. Asymmetric perception	Overall χ^2	Differences among profiles
Competence	3.932 (0.032)	3.340 (0.061)	3.388 (0.076)	98.327***	**1 > 2 = 3**
Benevolence	3.629 (0.048)	3.245 (0.056)	2.937 (0.068)	71.944***	**1 > 2 > 3**
Integrity	3.931 (0.038)	3.477 (0.057)	3.126 (0.065)	125.736***	**1 > 2 > 3**

Notes: ">" denotes differences between two profiles as to trust facets are significant at $p < 0.05$, whereas " = " denotes the differences above are not significant at $p < 0.05$; ***$p < 0.001$.

References

Ahmadian, M., & Lee, O. K. D. (2017). AI-based voice assistant systems: Evaluating from the interaction and trust perspectives. In *Proceeding of 23th americas conference on information systems*.

Asparouhov, T., & Muthen, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 329–341.

Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: An Interdisciplinary Journal, 23*(1), 20–31.

de Barcelos Silva, A., Gomes, M. M., da Costa, C. A., da Rosa Righi, R., Barbosa, J. L. V., Pessin, G., & Federizzi, G. (2020). Intelligent personal assistants: A systematic literature review. *Expert Systems with Applications, 147*, 113193.

Braun, M., Broy, N., Pfleging, B., & Alt, F. (2019). Visualizing natural language interaction for conversational in-vehicle information systems to minimize driver distraction. *Journal on Multimodal User Interfaces, 13*(2), 71–88.

Califf, C. B., Brooks, S., & Longstreet, P. (2020). Human-like and system-like trust in the sharing economy: The role of context and humanness. *Technological Forecasting and Social Change, 154*, 119968.

Chang, R. C. S., Lu, H. P., & Yang, P. (2018). Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in taiwan. *Computers in Human Behavior, 84*, 194–210.

Cheng, H. F., Yang, M. H., Chen, K. Y., & Chen, H. L. (2014). Measuring perceived EC ethics using a transaction-process-based approach: Scale development and validation. *Electronic Commerce Research and Applications, 13*(1), 1–12.

Cho, E., Molina, M. D., & Wang, J. (2019). The effects of modality, device, and task differences on perceived human likeness of voice-activated virtual assistants. *Cyberpsychology, Behavior, and Social Networking, 22*(8), 515–520.

Culley, K. E., & Madhavan, P. (2013). A note of caution regarding anthropomorphism in HCI agents. *Computers in Human Behavior, 29*(3), 577–579.

Demetis, D. S., & Lee, A. S. (2018). When humans using the IT artifact becomes IT using the human artifact. *Journal of the Association for Information Systems, 19*(10), 929–952.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170.

Duran, R. L. (1983). Communicative adaptability: A measure of social communicative competence. *Communication Quarterly, 31*(4), 320–326.

Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor's voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior, 90*, 357–362.

Evermann, J., & Tate, M. (2011). Fitting covariance models for theory generation. *Journal of the Association for Information Systems, 12*(9), 632–661.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299.

Foehr, J., & Germelmann, C. C. (2020). Alexa, can I trust you? Exploring consumer paths to trust in smart voice-interaction technologies. *Journal of the Association for Consumer Research, 5*(2), 181–205.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Gabriel, A. S., Daniels, M. A., Diefendorff, J. M., & Greguras, G. J. (2015). Emotional labor actors: A latent profile analysis of emotional labor strategies. *Journal of Applied Psychology, 100*(3), 863–879.
Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly, 27*(1), 51–90.
Gillath, O., Ai, T., Branicky, M., Keshmiri, S., Davison, R., & Spaulding, R. (2020). Attachment and trust in artificial intelligence. *Computers in Human Behavior, 115*, 106607.
Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018). Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction. In *Proceedings of the 26th European conference on information systems*.
Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior, 97*, 304–316.
Gursoy, D., Chi, O. H., Lu, L., & Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management, 49*, 157–169.
Haas, J. W., & Arnold, C. L. (1995). An examination of the role of listening in judgments of communication competence in co-workers. *International Journal of Business Communication, 32*(2), 123–139.
Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393–416.
Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*(5), 967–988.
Howard, M. C., & Hoffman, M. E. (2018). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods, 21*(4), 846–876.
Khatri, C., Venkatesh, A., Hedayatnia, B., Gabriel, R., Ram, A., & Prasad, R. (2018). Alexa prize—state of the art in conversational AI. *AI Magazine, 39*(3), 40–55.
Klaus, P., & Zaichkowsky, J. (2020). AI voice bots: A services marketing research agenda. *Journal of Services Marketing, 34*(3), 389–398.
de Kleijn, R., Wijnen, M., & Poletiek, F. (2019). The effect of context-dependent information and sentence constructions on perceived humanness of an agent in a Turing test. *Knowledge-Based Systems, 163*, 794–799.
Lankton, N. K., 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.
Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(1), 1–26.
Lee, K. M., Peng, W., Jin, S. A., & Yan, C. (2006). Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of Communication, 56*(4), 754–772.
Lortie, C. L., & Guitton, M. J. (2011). Judgment of the humanness of an interlocutor is in the eye of the beholder. *PloS One, 6*(9), Article e25085.
MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly, 35*(2), 293–334.
Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to the internal/external frame of reference model. *Structural Equation Modeling: An Interdisciplinary Journal, 16*, 1–35.
Martin, T. (2019). *73% don't trust conversational AI voice assistants*. Retrieved from https://www.mediapost.com/publications/article/336191/73-dont-trust-conversational-ai-voice-assistants.html.
Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734.
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.
McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance, 27*(3), 573–599.
Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437.
Meyer, J. P., Stanley, L. J., & Vandenberg, R. J. (2013). A person-centered approach to the study of commitment. *Human Resource Management Review, 23*(2), 190–202.
Morin, A. J., Morizot, J., Boudrias, J. S., & Madore, I. (2011). A multifoci person-centered perspective on workplace affective commitment: A latent profile/factor mixture analysis. *Organizational Research Methods, 14*, 58–90.
Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.
Niculescu, A., van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making social robots more attractive: The effects of voice pitch, humor and empathy. *International Journal of Social Robotics, 5*(2), 171–191.
Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture modeling using the three-step specification. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 439–454.
Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535–569.
Olson, P. (2019). *New report tackles tough questions on voice and AI*. Retrieved from https://about.ads.microsoft.com/en-us/blog/post/april-2019/new-report-tackles-tough-questions-on-voice-and-ai.
Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 616–639.
Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems, 25*(4), 145–182.
Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior, 109*, 106359.
Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*(5), 443–449.
Saffarizadeh, K., Boodraj, M., & Alashoor, T. M. (2017). Conversational assistants: Investigating privacy concerns, trust, and self-disclosure. In *Proceeding of 38th international conference on information systems*.
Santos, R., Abreu, J., Beça, P., Rodrigues, A., & Fernandes, S. (2020). Voice interaction on TV: Analysis of natural language interaction models and recommendations for voice user interfaces. *Multimedia Tools and Applications, 9*, 1–28.
Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems, 37*(3), 875–900.
Schuetz, S., & Venkatesh, V. (2020). The rise of human machines: How cognitive computing systems challenge assumptions of user-system interaction. *Journal of the Association for Information Systems, 21*(2), 460–482.
Schwartz, E. H. (2020). *Kids don't trust voice assistants*. https://voicebot.ai/2020/02/05/kids-dont-trust-voice-assistants-study.
Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research, 115*, 14–24.
Shin, M., Kim, S. J., & Biocca, F. (2019). The uncanny valley: No need for any further judgments when an avatar looks eerie. *Computers in Human Behavior, 94*, 100–109.
Sonpar, K., Handelman, J. M., & Dastmalchian, A. (2009). Implementing new institutional logics in pioneering organizations: The burden of justifying ethical appropriateness and trustworthiness. *Journal of Business Ethics, 90*(3), 345–359.
Specht, J., Luhmann, M., & Geiser, C. (2014). On the consistency of personality types across adulthood: Latent profile analyses in two large-scale panel studies. *Journal of Personality and Social Psychology, 107*(3), 540–556.
Srivastava, S. C., & Chandra, S. (2018). Social presence in virtual world collaboration: An uncertainty reduction perspective using a mixed methods approach. *MIS Quarterly, 42*(3), 779–804.
Strohmann, T., Siemon, D., & Robra-Bissantz, S. (2019). Designing virtual in-vehicle assistants: Design guidelines for creating a convincing user experience. *AIS Transactions on Human-Computer Interaction, 11*(2), 54–78.
Svenningsson, N., & Faraon, M. (2019). Artificial intelligence in conversational agents: A study of factors related to perceived humanness in chatbots. In *Proceedings of the 2019 2nd artificial intelligence and cloud computing conference*.
Tamagawa, R., Watson, C. I., Kuo, I. H., MacDonald, B. A., & Broadbent, E. (2011). The effects of synthesized voice accents on user perceptions of robots. *International Journal of Social Robotics, 3*(3), 253–262.
Torre, I., Goslin, J., & White, L. (2020). If your device could smile: People trust happy-sounding artificial agents more. *Computers in Human Behavior, 105*, 106215.
de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., et al. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied, 22*(3), 331–349.
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.
Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113–117.
Westerman, D., Cross, A. C., & Lindmark, P. G. (2019). I believe in a thing called bot: Perceptions of the humanness of "chatbots". *Communication Studies, 70*(3), 295–312.
Westerman, D., Edwards, A. P., Edwards, C., Luo, Z., & Spence, P. R. (2020). I-it, I-thou, I-robot: The perceived humanness of AI in human-machine communication. *Communication Studies, 71*, 393–408.
Wiese, E., & Weis, P. P. (2020). It matters to me if you are human-Examining categorical perception in human and nonhuman agents. *International Journal of Human-Computer Studies, 133*, 1–12.
Wise, A. F., & Hsiao, Y. T. (2019). Self-regulation in online discussions: Aligning data streams to investigate relationships between speaking, listening, and task conditions. *Computers in Human Behavior, 96*, 273–284.
Woo, S. E., Jebb, A. T., Tay, L., & Parrigon, S. (2018). Putting the "person" in the center: Review and synthesis of person-centered approaches and methods in organizational science. *Organizational Research Methods, 21*(4), 814–845.
Xie, H., Prybutok, G., Peng, X., & Prybutok, V. (2020). Determinants of trust in health information technology: An empirical investigation in the context of an online clinic appointment system. *International Journal of Human-Computer Interaction, 36*(12), 1095–1109.
Xu, K. (2019). First encounter with robot Alpha: How individual differences interact with vocal and kinetic cues in users' social responses. *New Media & Society, 21*(11–12), 2522–2547.
Yang, S., Jiang, H., Yao, J., Chen, Y., & Wei, J. (2018). Perceived values on mobile GMS continuance: A perspective from perceived integration and interactivity. *Computers in Human Behavior, 89*, 16–26.
Zyphur, M. J. (2009). When mindsets collide: Switching analytical mindsets to advance organization science. *Academy of Management Review, 34*(4), 677–688.