International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Understanding and Enhancing Human-Voice Assistant Collaboration: Developing and Validating a Collaboration Intensity Scale

Sarah Zabel, Marah Blaurock, Marion Büttgen & Siegmar Otto

To cite this article: Sarah Zabel, Marah Blaurock, Marion Büttgen & Siegmar Otto (24 Apr 2025): Understanding and Enhancing Human-Voice Assistant Collaboration: Developing and Validating a Collaboration Intensity Scale, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2025.2491790

To link to this article: https://doi.org/10.1080/10447318.2025.2491790

Published online: 24 Apr 2025.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2025.2491790

Understanding and Enhancing Human-Voice Assistant Collaboration: Developing and Validating a Collaboration Intensity Scale

Sarah Zabel[a], Marah Blaurock[b,c], Marion Büttgen[d], and Siegmar Otto[a]

[a]Department of Sustainable Development and Change, University of Hohenheim, Stuttgart, Germany; [b]Department of Service Management, Ingolstadt School of Management, Catholic University Eichstätt-Ingolstadt, Germany; [c]Institute for Applied AI, Stuttgart Media University, Stuttgart, Germany; [d]Department of Corporate Management, University of Hohenheim, Stuttgart, Germany

ABSTRACT

The collaboration between humans and voice assistants (VAs) is increasingly relevant in both professional and everyday contexts. However, valid measures to capture users' behavioral contributions to this collaboration are lacking. This study introduces a behavior-based scale to measure collaboration intensity—the extent to which users engage effectively and productively with VAs in collaborative tasks. Two surveys ($N_1 = 319$; $N_2 = 321$) and Item Response Theory were used to develop and validate a 30-item scale with high reliability (EAP = .79/.88) and good item fit, alongside a 12-item short version (EAP = .77). Construct validation showed positive correlations with technology acceptance and negative correlations with commitment to informational self-determination. The scale also demonstrated predictive validity and an association with extraversion. The Collaboration Intensity Scale offers researchers and practitioners a tool to assess and enhance human–VA collaboration by identifying behavioral patterns and informing the design of more effective VA features.

1. Introduction

Technologies enabled by artificial intelligence (AI) are becoming increasingly prevalent, if not ubiquitous in various areas of life, ranging from the workplace to everyday life (Terzopoulos & Satratzemi, 2020). In both professional and consumer contexts, collaborative intelligence systems (i.e., collaborative AI-based technologies) help humans achieve their or their organization's goals and enhance efficiency (Epstein, 2015; Henkel et al., 2020; Maedche et al., 2019; Wilson & Daugherty, 2018). Examples of such collaborative intelligence systems include conversational agents (CAs)—systems that simulate interactions with humans in written or spoken language—such as ChatGPT, or voice assistants like Apple's Siri, or Amazon's Alexa (Mariani et al., 2023; Radziwill & Benton, 2017).

Supported by CAs, users can organize their shopping, make appointments, cook meals, optimize their personal music playlists, and find their way to various destinations (Bokolo Jnr, 2024; Lee et al., 2023; PWC, 2018). At work, CAs help users improve customer support, draft legal documents, or prepare emails (Chui et al., 2022), and human engagement with these systems is evolving from mere usage to collaboration (Blaurock et al., 2024). The CA helps users achieve their goals, while the users' behavior and feedback, in turn, helps the CA to adapt its predictive model and offer more accurate recommendations in the future (Guo, 2020). Existing studies have predominantly focused on a system's characteristics such as its design, as well as privacy-related aspects of CAs as a driver of user acceptance, usage intentions (e.g., Cheng & Jiang, 2020; Dutsinma et al., 2022; Mariani et al., 2023), or users' trust towards the system (e.g., Fernandes & Oliveira, 2021; Kasilingam, 2020; Leschanowsky et al., 2024). Additionally, research has focused on fair and inclusive AI systems by examining biases in the design of AI in general and CAs in particular (e.g., Zabel & Otto, 2021, 2024), for instance regarding the reinforcement of gender stereotypes (e.g., a subordinate attitude, Hwang et al., 2019), and discrepancies in voice recognition accuracy for different speaker groups (e.g., non-native speakers, speakers of regional accents, Feng et al., 2024). This focus on the design of fair, inclusive, and transparent CAs is essential for maximizing their collaborative potential, providing the best opportunities for user engagement. Yet, the effectiveness of a human-CA-collaboration relies not only on the system's features, but also on the users' contributions to facilitating and enhancing collaboration, characterized by joint effort and an integration of the respective abilities to complete a task (Maedche et al., 2019; Mattessich & Monsey, 1992; Sowa et al., 2021). Consequently, Mariani et al. (2023) emphasized the importance of studying users' inherent characteristics, which facilitate CA adoption and human-CA collaboration (Hering et al., 2024; Zabel et al., in press). Attig et al. (2017) argued for more extensive research on interindividual differences in human-technology interaction to understand user diversity and increase user-technology fit. Yet, there remains a gap

CONTACT Sarah Zabel sarah_zabel@uni-hohenheim.de Department of Sustainable Development and Change, University of Hohenheim, 70599 Stuttgart, Germany

in understanding how users contribute to a successful collaboration with a CA through their behavior. Addressing this gap is essential, as CAs are not static, but are actively shaped by users through their engagement and feedback.

The present study addresses this gap by developing and validating a scale to measure human-CA collaboration intensity—the extent to which a user engages effectively and productively with CAs. Our primary contribution lies in the introduction of a novel scale to assess users' contributions to the collaboration with a voice assistant (VA)—a voice-based CA—focusing on their actual collaboration behaviors rather than perceptions or attitudes. By applying Item Response Theory (IRT), the scale allows for the precise assessment of individuals based on the intensity of collaboration with VAs (see also Bond & Fox, 2013). Furthermore, the combination of behavior-based measurement and IRT, which facilitates item-level analyses, enables the identification of areas where collaboration might still be difficult for users and, thus, informs improvements in VA design and alignment with user needs. This includes the development and marketing of more personalized VAs and voice-based services, which cater to diverse user needs and preferences (Cerezo et al., 2025). Additionally, it will be possible to extend the scale in the future to take account of dynamic developments in the field of VAs. For more economical testing, we also proposed a short scale with 12 items. Finally, unlike existing scales measuring technology acceptance and related constructs (Neyer et al., 2012), our scale emphasizes collaboration behavior, assessing the user's contribution to the collaborative process, thereby advancing the theoretical understanding of human-VA collaboration.

2. Theoretical background

CAs are AI-enabled agents that work via natural language processing, allowing users to interact with them via written or spoken language (Mariani et al., 2023). They perform a range of tasks, including searching for information (e.g., *daily news*), ordering food or consumer goods, translating language, and even managing other smart devices (Hoy, 2018; OpenAI, 2023). They differ from other technologies because virtually no technical knowledge is needed to use them, and they offer a convenient way to interact and complete tasks, similar to engaging in a dialog with a human counterpart (Guzman, 2019; McLean et al., 2021). Of all CAs, the interaction modality of spoken language enables voice assistants (VA) to interact in the most natural and intuitive way (Rzepka et al., 2022). Furthermore, VAs are particularly suited to performing a variety of tasks while the user is busy with other tasks (e.g., sending dictated messages while driving). It is thus unsurprising that VAs are becoming increasingly popular (Deloitte, 2019; Maedche et al., 2019)—in 2022, 47% of internet users controlled their electronic devices at least occasionally via voice command, with even higher usage among those aged 16–29 (Bitkom, 2022) and the global number of AI-based voice assistants in use is estimated to reach eight billion in 2024 (Statista, 2024).

2.1. Existing scales and measurement

Researchers have explored how the design of CAs in general and VAs in particular, as well as user-related aspects, such as attitudes or perceptions, influence system adoption (Mariani et al., 2023), and scales have been developed for the particular context of human-CA interaction. Table 1 presents an overview of these measures and indicates what the scale measures and through whose experience (i.e., unit of measurement; MacKenzie et al., 2011). We also indicate which type of human response to interactions with CAs the scales measure leveraging the well-established Affect-Cognition-Behavior Model (ABC; Breckler, 1984). Of the six scales, three focus on VAs, with two of them measuring

Table 1. Overview of existing scales related to human-CA interaction.

Reference	Scale name	Aim	Unit of measurement	ABC category
Chen et al. (2022)	AI chatbot service quality (AICSQ)	Theoretical classification and measurement development for AICSQ, which provides a framework and benchmark for evaluating, designing, and improving AI chatbot service quality for practitioners.	Users' perceptions of CA quality	C
Klein et al. (2020)	UX with VA scale	Evaluating VA quality through user experience (UX) to evaluate and improve VAs.	Users' perceptions of CA interaction experience	C
Zwakman et al. (2020)	Voice Usability Scale (VUS)	Construct and validate a scale for subjective usability evaluation of VAs.	Users' perceptions of CA usability	C
Borsci et al. (2022, 2023)	The Chatbot Usability Scale	Providing a toolkit to help chatbot designers consider the needs of the end-users in the design process to achieve user satisfaction.	Users' perceptions of CA usability	C
Banks (2019)	Perceived Moral Agency (PMA)	Measurement of human interlocutors' perceptions of social machines (e.g., CAs) as moral agents through perceived moral agency.	Users' perceptions of CA as a moral agent	C
Wienrich and Carolus (2021)	Conversational Agent Literacy Scale (CAL)	Expand literacy concepts to account for AI-related specifics in the context of voice-based conversational agents; assessment of conceptualizations and competencies about conversational agents.	Users' knowledge of CA functionalities	C
Our scale	Human-VA Collaboration Intensity	Establishing a behavioral scale to assess human-VA collaboration Intensity based on IRT.	Users' behavior when collaborating with voice assistant	B

Note. ABC: Affect-Behavior-Cognition (Breckler, 1984); IRT: Item Response Theory.

usability (Klein et al., 2020; Zwakman et al., 2020) and one assessing literacy (Wienrich & Carolus, 2021).

In studying the existing scales, it becomes apparent that they predominantly focus on measuring users' perceptions of the CA's quality (Chen et al., 2022), the CA's usability (Borsci et al., 2022; Zwakman et al., 2020), or the user's interaction experience (Klein et al., 2020), all of which adhere to the category of cognitive responses to environmental stimuli rather than affective or behavioral responses (Breckler, 1984). Aside from the Conversational Agent Literacy Scale (CAL), which measures knowledge about CA functionalities (Wienrich & Carolus, 2021), no other scale measures users' contributions to the interaction process. The CAL scale focuses on conceptualizations and competencies about CAs, and it incorporates—amongst others—users' knowledge about and skills in interacting with CAs. The CAL scale thus provides insights into users' cognitive understanding of CAs, which is a predictor of the literate use of a CA. However, it does not focus on actual collaboration behavior. Therefore, we propose a scale which measures human-VA collaboration intensity via actual engagement with the VA, to complement existing scales which focus on cognitive responses to or evaluations of the system.

Furthermore, most of these measures have some common shortcomings: They mainly assess self-reported perceptions and attitudes (e.g., Loyd & Gressard, 1984), which cannot keep up with self-reported behavior in terms of validity, as answering questions about attitudes requires a high degree of reflection among the participants and often results in common method bias (e.g., Otto et al., 2018). Despite the advantages of behavioral measures as compared with classical measures of attitude, only a few studies have focused on user behavior when interacting with voice assistants. For example, Purington et al. (2017) investigated user interaction sociability with Amazon Echo, and Sciuto et al. (2018) explored households' integration of voice assistants by analyzing the users' requests. A scale that examines past behavior as a reflection of users' contribution to the human-VA collaboration would allow researchers and practitioners to rapidly collect valid data on human-VA collaboration intensity across CA brands and with a variety of users across the globe.

2.2. Human-voice assistant collaboration

In many areas, the partnership between users and VAs can increasingly be viewed as a collaboration. Human-VA collaboration—in contrast to interaction, which is a more general term that refers to two or more entities actively (socially) engaging with each other—is defined as a human and a digital device or robot working together on a joint task or goal (Fong et al., 2003; Merriam-Webster, 2023; Sowa et al., 2021). When collaborating, two entities work in concert, and individual abilities are integrated to complete a joint task (Maedche et al., 2019; Mattessich & Monsey, 1992; Sowa et al., 2021). Thus, collaboration strives for a mutually beneficial interworking in symbiosis (Maedche et al., 2019; Paschen et al., 2020).

AI-based systems that partner up with a user to achieve the user's goals are called collaborative intelligence systems (Epstein, 2015). Such systems are intended to engage in different tasks with users as active collaborators (Epstein, 2015). For example, VAs collaborate with users in developing musical ambience for different occasions. The voice assistant makes suggestions on the basis of an analysis of past user data while the user provides feedback to obtain the ideal playlist (Guo, 2020). It is only through this collaboration and feedback that the voice assistant can learn leading to improved collaboration results over time (e.g., Guo, 2020).

There are numerous examples of VAs collaborating on tasks with humans in the work context as well. One example is Alexa for Business, which serves as a personal assistant, performing tasks, such as scheduling meetings, booking conference rooms, retrieving reports or slides, and even enabling users to make calls and check security camera feeds (Forbes Technology Council, 2020). In the logistics sector, voice assistants can streamline delivery processes by supporting drivers through apps built into their vehicles or scanners (StageZero Technologies, 2022). With their natural language processing capabilities, VAs are designed to be a more human-like technology than predecessor technologies (Guzman, 2019). They engage users in a more interactive manner, resembling the way humans interact with each other and thus fostering social connections and collaboration (van Pinxteren et al., 2023). Thus, while there are a multitude of scales for assessing users' propensity to accept and use technology (see Attig et al., 2017 for an overview), the collaborative nature of VAs and similar technologies raises questions about the applicability of these scales.

2.3. Collaboration intensity

Our Collaboration Intensity construct relates to established theoretical technology acceptance models, namely the extended Technology Acceptance Model (TAM2) (Venkatesh & Davis, 2000) and the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003) and extant conceptual work on human-AI collaboration (Blaurock et al., 2024; Epstein, 2015). The technology acceptance models already provide valuable foundations for understanding users' interactions with information technology such as VAs as they identify critical antecedents including perceived usefulness and performance expectancy, instrumental factors (e.g., ease of use, job relevance, effort expectancy, facilitating conditions), and social influence (e.g., subjective norm, image) as well as important moderators (e.g., age, gender, previous experience) of technology use (Venkatesh et al., 2003; Venkatesh & Davis, 2000). Our work relates to these theoretical models as we offer additional insights into their behavioral backend. While TAM and UTAUT focus on the mere intention and subsequent actual use of technology, Collaboration Intensity allows for fine grained analysis of users' level of engagement with a collaboration technology. Moreover, our construct draws from conceptual collaborative intelligence models that focus specifically on collaboration rather than interaction with interactive technology (i.e., collaborative intelligence systems such as CA, VA, social robots) (Epstein, 2015; Blaurock et al., 2024). These models show that for effective collaboration, users and systems

are interdependent, enhancing (or compensating for) each other's abilities and specify these abilities as detailed below. This interdependency requires both parties to contribute to the collaboration, leveraging their strengths to achieve the best outcome. Thus, the success of a human-technology collaboration hinges not only on the system's features but on the propensity of humans to engage with a technology, and on their active collaboration (Epstein, 2015). We term this human contribution to an enabling and enhancing collaboration with collaborative intelligence systems *Collaboration Intensity*. It refers to *the extent to which an individual is engaging effectively and productively with a collaborative intelligence system (e.g., VAs) in pursuit of a joint task or goal.*

For our scale, we focus on human-VA collaboration and draw from research on collaborative intelligence (Blaurock et al., 2024; Epstein, 2015) to understand and structure the different contributions of users and VAs in the collaboration. Recent research has identified five design features of collaborative intelligence systems that are related to favorable outcomes such as perceived service improvement, namely engagement, transparency, process control, outcome control, and reciprocal strength enhancement (Blaurock et al., 2024). To address the interdependent nature of the collaboration process, we adapt this concept from the system's prerequisites to the user's contribution to the collaboration process. Our newly developed scale assesses human behavior to infer the human contribution to the collaboration, which can fall into different behavioral categories: (a) Active engagement describes the human agent actively approaching the voice assistant. Epstein (2015) emphasizes that co-creation of outcomes is essential for successful human-AI collaboration. Human users can actively engage systems to participate in the collaboration process (e.g., the user actively asks for suggestions). (b) Transparency refers to the user making themself transparent to the voice assistant, that is, providing data and, thereby, enabling the voice assistant to perform tasks. Providing such information (e.g., granting access to contacts or location) is vital for the system to be able to contribute to the collaboration process (Jiang et al., 2024) (c) Reciprocal strength enhancement consists of two parts: either the user strengthens the system's performance by, for example, giving feedback, or the system enhances the user's strengths by the user leveraging the system to take over tasks it can accomplish more efficiently than they (e.g., because the system can support conversations in foreign languages, the users scope of action, for example, to learn something new, increases significantly). Reciprocal strength enhancement is essential for continuous improvement of human and machine for effective and productive collaboration (Guo, 2020; Wilson & Daugherty, 2018). (d) Process control describes the behavior of altering an interaction process (e.g., deciding how to use an application in the collaboration process with the system such as skipping a movie part). And finally, (e) outcome control refers to behaviors related to providing outcome control over tasks or not (e.g., letting the VA send messages). Being able to exhibit control in human-technology interactions is essential for user trust (Rohden & Espartel, 2024) and satisfaction (Fan et al., 2022). Because of their close conceptual relation and limited number of potential control behaviors in our context, we merged the subdimensions of process control and outcome control into one control dimension.

2.3.1. Related constructs

2.3.1.1. Technology acceptance. The strongest conceptual similarity between the collaboration intensity and known constructs was expected for technology acceptance. Technology acceptance refers to the extent to which people are generally open and enthusiastic about new technologies, thus reflecting a type of attitude toward technology in general (Neyer et al., 2012). With regard to VAs, enthusiasm and openness about new technologies could make it easier to enter into a collaboration and try new input (e.g., prompts), thereby paving the way for positive collaboration experiences and an intensive collaboration. It is therefore likely that this overall acceptance of new technologies is also related to the more specific collaboration intensity. However, there are certain differences between these two constructs: Collaboration intensity focuses specifically on VAs, rather than technology in general. Thereby, it is possible to obtain a more precise picture of how users collaborate with such systems, and to study the determinants and constructs that are related to the collaboration intensity (e.g., personality traits). Furthermore, technology acceptance focuses primarily on users' willingness to adopt and use a technology (Venkatesh et al., 2003) on the basis of their perception of it and on other external and internal factors. In contrast, collaboration intensity assesses users' behavior when collaborating with a VA, emphasizing the interdependent nature of the user-VA interaction.

Therefore, on the basis of the abovementioned points, we hypothesized:

H1: *Collaboration intensity is positively correlated with technology acceptance (construct validity).*

2.3.1.2. Commitment to informational self-determination. In order to collaborate with VAs, people must be willing to disclose personal information. Therefore, having a low commitment to informational self-determination (i.e., having a positive attitude toward disclosing personal data) is likely to be related to collaboration intensity (see also Easwara Moorthy & Vu, 2015). However, due to the conceptual dissimilarity between collaboration intensity and the commitment to informational self-determination, we expected the correlation to be weak. Thus, we hypothesized:

H2: *The human-VA collaboration intensity is weakly negatively correlated with commitment to informational self-determination (construct validity).*

2.3.2. Determinants of collaboration intensity

Personality traits predict individual differences in attitudes and behavior in various domains (Schepman & Rodway, 2023). Research on technology acceptance and use has

shown that individuals who are high in openness tend to give higher ratings to the perceived usability of technologies (Kortum & Oswald, 2018). Furthermore, a person with high openness to experience is described as being innovative and unconventional (Ashton & Lee, 2008). Thus, we expected that people higher in openness would be more likely to adopt and collaborate with voice assistants, and thus, openness and collaboration intensity were expected to be correlated.

Another relevant personality trait is extraversion. It was found to be negatively correlated with technology use (Barnett et al., 2015), and attitude towards AI (Bergdahl et al., 2023; Park & Woo, 2022). However, we expected it to be positively related to collaboration intensity. Extraversion is related to engaging in social activities (e.g., Argyle & Lu, 1990). Thus, individuals high in extraversion may enjoy collaborating with a CA and especially a VA, which comes close to a human conversational partner due to its communication in natural language (Guzman, 2019). By examining the relationship between collaboration intensity and extraversion, our aim is to assess the uniqueness of the collaboration intensity construct, demonstrating that it differs from technology acceptance and use. We hypothesized:

H3: *(a) Openness to experience and extraversion (b) are positively related to collaboration intensity (construct validity).*

2.3.3. Collaboration intensity in voice assistant power users

Finally, we aimed to test the appropriateness of the collaboration intensity scale for identifying power users (i.e., users who use more than one voice assistant or users who use voice assistants several times a day). We expected that collaboration intensity would be positively related to (a) the number of voice assistants used and (b) the frequency of voice assistant use. Furthermore, we expected collaboration intensity to explain additional variance when technology acceptance was included as another predictor in the model. Thus, we hypothesized:

H4: *Collaboration intensity predicts (a) the frequency of voice assistant use and (b) the use of different voice assistants (predictive validity).*

2.4. Measurement approach

We measure collaboration intensity with a behavioral scale, that is, by assessing whether users perform certain behaviors when using a voice assistant. Using questions about performed behaviors comes with the advantage that answers are less prone to response biases (e.g., Otto et al., 2018) and can be answered without intensive self-reflection (i.e., more quickly). Furthermore, behavioral items can be ordered by the difficulty of the behavior, with difficulty reflecting the costs (not merely financial, but also in terms of personal resources such as time, trust, or effort) associated with the behavior (see also Brügger et al., 2011; Kaiser et al., 2010). For instance, the item "I let my voice assistant pay for me" may be considered more difficult than "I ask my voice assistant for the latest news" because, among other things, the VA would gain access to extensive personal data in the former, but not in the latter case. Ordering the behaviors according to their difficulties has two main advantages. First, it makes it possible to identify behaviors that are perceived as particularly difficult, which can be considered when designing VAs. Second, based on an individual's responses to these items, it is possible to derive their collaboration intensity: The higher the collaboration intensity, the more collaborative behaviors (with a high level of difficulty) are performed.

We analyze the data based on IRT, which is particularly suitable for handling items with varying levels of difficulty (i.e., collaboration behaviors with different levels of necessary effort, time, etc.). Compared to Classical Test Theory (CTT), IRT provides several advantages for scale development and validation. While CTT primarily focuses on test-level information, IRT estimates both item and person parameters on the same scale (e.g., Bond & Fox, 2013), allowing for precise measurement of the latent trait, in our case individual collaboration intensity. The individual scores are independent of the specific test items used, thus providing a stable measure of an individual's collaboration intensity, even if items change. This stability makes IRT especially suitable for adaptive testing scenarios in which the selection of items is based on an individual's responses to previous items, which allows for a precise estimation of an individual's collaboration intensity based on a smaller number of items. Furthermore, it offers the possibility of including new items to the scale (cf. Bond & Fox, 2013), which is especially beneficial as VAs and their functionalities dynamically evolve and users' contributions to the collaboration may change. IRT has been successfully used in behavioral assessments (Kaiser & Wilson, 2004) and has proven effective in Human-Computer Interaction research (e.g., Berkman & Şahin, 2024; Cai et al., 2024). With regards to our scale, IRT enables detailed item analysis, helping to identify which behaviors are perceived as more difficult (Bond & Fox, 2013). This information can then be used to align the design of VAs more closely with user needs.

3. Scale development and validation process

Figure 2 illustrates the process and content of our scale development and validation process. After consulting previously developed scales to ensure the value added of our scale, we developed 70 behavioral items to measure collaboration intensity following Kaiser et al.'s (2010) recommendations. We then conducted two validation studies with online convenience and online representative samples. In the first study, we used IRT to internally validate our newly developed items by assessing item fit (especially item infit statistics, see Linacre, 2002) and difficulty patterns to ensure that the selected items cover the full range of collaboration intensity and, thus, allow for a precise differentiation across varying levels of intensity. We ensured that items which exhibited poor fit or redundancy were excluded (see also Section 5.3). We selected 30 items for the final scale. We also assessed construct and predictive validity. In a final step, we conducted a second study to validate the scale in a

more heterogeneous sample internally, and we used additional constructs for the external validation. Finally, we identified suitable items for a short scale. Our aim was to obtain a scale that allows for economic testing, while still demonstrating satisfactory reliability. We followed the same item reduction procedure as described above, resulting in a 12-item scale.

4. Item development

Our aim was to develop a set of items that could be used to measure collaborative behavior with voice assistants to infer users' collaboration intensity. To do so, we followed the procedure described by Kaiser et al. (2010). The items were developed to cover a wide range of voice assistant functions and varied only in their difficulty. They were grouped according to the categories described in Section 2.3 (cf. Blaurock et al., 2024).

The item generation proceeded as follows: First, three researchers independently engaged in brainstorming to generate a diverse set of items that reflect actual behaviors of users when collaborating with voice assistants. Next, they convened to jointly discuss and compare the items and engaged in test interactions with the voice assistants *Google Assistant* and *Siri* (two voice assistants with great coverage in Germany at the time of the study [June 2021], see Splendid Research, 2019) with the aim of discovering additional collaboration functions. Furthermore, insights into the functions commonly used by voice assistant users were gathered by leveraging the personal networks of the researchers. Finally, user statistics with regard to specific functions were taken into account to develop items (Statista, 2021). The aim of using multiple sources for the item generation was to cover all identifiable voice assistant functions at the time of the study. The behaviors covered include basic interactions (e.g., setting a timer), task delegation (e.g., sending messages), data-sharing (e.g., speech analysis; granting access to one's contacts), and decision-making support (e.g., asking the voice assistant multiple times before making a decision). The researchers discussed the resulting collection of 70 items and aligned them with the categories identified by Blaurock et al. (2024)—active engagement, transparency, reciprocal strength enhancement, process and outcome control (see also Section 2.3).

To ensure that the items cover a broad range of collaboration behaviors, they were designed to reflect both simple and more complex or demanding collaboration. Some items were based on basic functions such as setting a timer, while others captured more advanced behaviors like using the voice assistant for financial transactions, or asking it several times before making a decision. This approach allows us to differentiate between users with minimal engagement and those who frequently and intensively collaborate with their voice assistant.

The matching process was conducted systematically: The three researchers who generated the items independently categorized each developed item. Their individual categorizations were then compared and discussed with two additional researchers who had not been involved in the item generation process. This group discussion ensured that each item was appropriately assigned and that items that lacked face validity, or were redundant, were removed. The first study was conducted with these 70 items, with the aim of selecting about 30 items for the final scale.

5. Study 1

When designing and conducting the study, we ensured that it is in accordance with the Declaration of Helsinki (World Medical Association [WMA], 2013) and the American Psychological Association's Ethical Principles (American Psychological Association [APA], 2017). As our study does not involve any procedures that would induce stress, nor high-risk participants, ethical approval was not required according to the German Research Foundation (Deutsche Forschungsgemeinschaft, 2023).

5.1. Sample and procedure

Data were collected with an online questionnaire on the SoSci Survey platform (https://soscisurvey.de) from June 11 2021 to July 2 2021. Participants were recruited by three researchers. At the beginning of the survey, all participants were informed about the purpose of the research, the voluntary nature of their participation, data anonymity, and their right to withdraw from the study at any time without providing a reason. No personally identifyable or sensitive data were collected. Participants then indicated their agreement to these terms and consented to participate in the study. Besides personal contacts, the ad hoc sample was contacted via a university mailing list (a Master's student distribution list), personal posts on social media platforms (LinkedIn, Facebook, Instagram, and Twitter [later renamed X]), and shares in facebook interest groups related to AI, Alexa, and Amazon Echo, as well as local community bulletin boards, and a notice posted in stores nearby. No incentives were

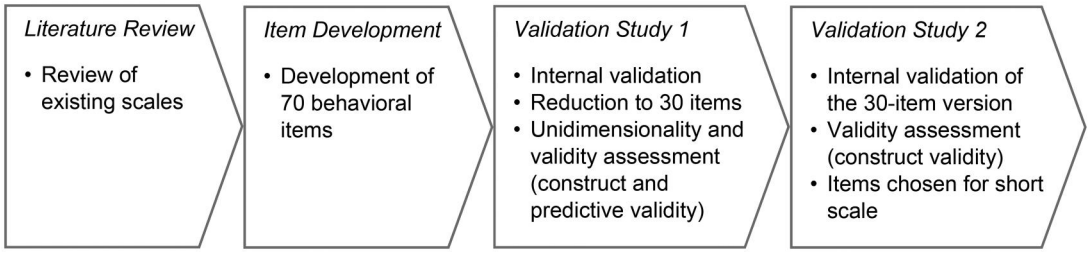

Figure 2. Scale development and validation process.

offered for participation. The sample was restricted to participants who had used a voice assistant during the last 2 years. A total of 459 participants responded to the survey. We removed data from participants who answered less than 75% of the questions, did not complete the Collaboration Intensity Scale, responded too quickly (i.e., more than twice as fast as the median response time; see Leiner, 2019), or did not pass the attention check, which consisted of participants having to select a specific answer category for one question. The final sample consisted of $N = 319$ participants. Their ages ranged from 18 to 74 years, with a mean age of 29.74 years ($SD = 11.20$); 58.93% were women. As the highest educational qualification, 152 individuals stated that they had a university degree or higher, 40 vocational training, 97 a high school diploma, and 15 either a secondary school diploma or a lower secondary school diploma.

5.2. Measures

5.2.1. Collaboration intensity

Our main variable of interest—collaboration intensity—was measured using our newly developed scale which assesses the extent to which users engage in collaborative behaviors with their voice assistant. This construct is based on the Collaborative Intelligence features (Blaurock et al., 2024) and reflects varying levels of active engagement, transparency, reciprocal strength enhancement, as well as process and outcome control. The response format of the items was either dichotomous (yes/no) or consisted of a 5-point scale with levels ranging from 1 (*never*) to 5 (*very often/always*) or from 1 (*strongly disagree*) to 5 (*strongly agree*). All had an additional category (not specified), which was to be selected if the person could not provide any information on an item because their voice assistant did not include the specific function or users do not perform the task (e.g., if someone does not watch movies or series, they cannot use their voice assistant for controlling the device). Items were coded in such that higher values indicated greater collaboration intensity. Some items were reverse-coded, as indicated in Appendix Table A1. The scores were then dichotomized, which is a prerequisite for the Rasch analysis, so that response categories three to five were recoded as 1 (i.e., the behavior is performed), and response categories 1 and 2 were recoded as 0 (i.e., the behavior is not performed). While dichotomization has been discussed as leading to a loss of information (MacCallum et al., 2002), it can also reduce measurement error (DeCoster et al., 2009). With regards to behavioral scales, researchers found that participants had difficulties answering frequency questions consistently, thus, dichotomization also serves to reduce measurement error in behavioral scales (Kaiser & Lange, 2021). Alternative methods such as the Partial Credit Model were considered less suitable for our scale's intended application, as they may introduce greater response inconsistencies. Furthermore, parts of our items were already in a dichotomous format in the questionnaire, because they reflected whether a behavior was performed or not. The dichotomous Rasch Model was thus chosen as the best fit to capture collaboration intensity based on the occurrence of behaviors.

5.2.2. Technology acceptance

To assess construct validity, four items from a short scale developed by Neyer et al. (2012) were used to assess technology acceptance using a 5-point scale ranging from 1 *(do not agree at all)* to 5 *(fully agree)*. Individuals rated their level of agreement with four statements regarding their attitude toward new technology (e.g., "I am very curious about new technological developments"). The internal consistency was measured with Cronbach's α ($\alpha = .87$), $M = 3.61$ ($SD = 0.87$).

5.3. Results

Our analyses were performed using R, version 4.1.2 (R Core Team, 2020, 2021). Besides the TAM package which was used to calibrate the Rasch model for all 70 items (Robitzsch et al., 2020), we used the following R packages for data management and statistical analyses (e.g., reliability estimates for the other scales, or the calculation of beat coefficients): psych (Revelle, 2021), lmtest (Zeileis & Hothorn, 2002), dplyr (Wickham et al., 2021), car (Fox & Weisberg, 2019).

Based on item statistics, we selected 30 items for the final scale. We focused especially on MS infit, which is ideally around 1 (values between 0.8 and 1.2 are considered uncritical, see Wright & Linacre, 1994 for reference values), and selected items with a broad difficulty distribution for the final scale. When there were items with similar difficulty, our objective was to streamline the selection, focusing on the infit and the content of the items to cover different aspects of the collaboration with voice assistants. For instance, in the initial item pool, two items involved dictating messages to the voice assistant—one sent automatically, the other manually. Since both had similar difficulty, only one was included in the final scale to reduce redundancy. Additionally, we evaluated Differential Item Functioning to ensure no substantial variation in how items discriminated across different groups—in our case, different genders (Holland & Wainer, 2012).

For the remaining 30 items, we recalibrated the scale, re-examining the fit statistics and the distribution of item difficulties, which is presented in Table 2. Some behaviors were rarely performed, for instance, most participants did not have their VA pay for them, whereas other behaviors were more common (e.g., giving the voice assistant new input if it misunderstands the user). Overall, there were more difficult than easy items. We further tested for unidimensionality by conducting a principal components analysis (PCA) of the standardized residuals of the Rasch model (Mylona et al., 2021) to examine if there are further factors explaining a substantial portion of variance in the data. Unidimensionality can be assumed, as the PCA revealed that the first component explained less than 15% of the variance in the residuals, and the variance explained by subsequent components decreased gradually (Fisher, 2007). The scale's expected a posteriori (EAP) reliability was good, EAP = .79 (see also Bühner, 2011 for reference values).

Additionally, we examined the correlation between the Collaboration Intensity scale with gender and age. Both correlations were not statistically significant ($r = -.06$, $p = 0.27$ for gender and $r = -.02$, $p = 0.66$ for age).

Table 2. Item difficulties for the 30-item scale (studies 1 and 2).

Number	Text	Category	δ_1	p_1	δ_2	p_2
1	I have my voice assistant pay for me.	AE	6.26	0.00	4.01	0.02
2	I call a taxi with the help of my voice assistant.	AE	5.14	0.01	3.13	0.04
3	I use my voice assistant to get food delivered.	AE	4.84	0.01	3.12	0.04
4	I schedule appointments for my family with my voice assistant.	RS	3.44	0.03	2.91	0.05
5	I use my voice assistant to engage in conversations in a foreign language.	RS	3.12	0.04	2.39	0.09
6	I have my voice assistant suggest leisure activities.	AE	2.91	0.05	1.99	0.12
7	I use my voice assistant for work.	RS	2.51	0.08	2.29	0.10
8	I have instructed my voice assistant to address me by my nickname.	T	2.42	0.08	2.55	0.07
9	I use my voice assistant while watching a series, film, etc. (e.g., start, stop, fast-forward, and rewind).	C	2.19	0.10	2.00	0.12
10	I have my voice assistant solve computing tasks.	AE	2.04	0.12	1.46	0.19
11	I have my voice assistant suggest gas stations/charging stations to me.	AE	2.02	0.12	1.74	0.15
12	I use my voice assistant to suggest restaurant alternatives nearby.	AE	1.99	0.12	1.39	0.20
13	I dictate messages to my voice assistant and allow it to send them.	C	1.93	0.13	1.40	0.20
14	I dictate messages to my voice assistant and send them myself afterwards.	C	1.85	0.14	1.48	0.19
15	I have my voice assistant read out my messages (e.g., WhatsApp, SMS, e-mails).	AE	1.78	0.15	1.61	0.17
16	Before I ask anyone else, I ask my voice assistant.	AE	1.71	0.16	1.51	0.18
17	I ask my voice assistant several times before making a decision.	C	1.62	0.17	1.85	0.14
18	The way I treat my voice assistant is similar to the way I treat a human counterpart (e.g., talk, interact).	–	1.08	0.25	1.02	0.27
19	I allow my voice assistant to access all my personal data.	T	0.48	0.38	1.14	0.24
20	I ask my voice assistant when I don't know something and thus learn new things.	RS	0.30	0.43	−0.75	0.68
21	If I want to research something, I use my voice assistant to do so.	AE	0.29	0.43	−0.34	0.58
22	When I hear a song I like, I use my voice assistant to find out the name of the song.	RS	0.21	0.45	−0.55	0.63
23	I use my voice assistant for navigation.	RS	0.14	0.47	0.14	0.47
24	When I'm driving a car, I use my voice assistant to call someone.	AE	0.13	0.47	0.68	0.34
25	I grant my voice assistant access to my contacts.	T	−0.30	0.57	0.45	0.39
26	I have my voice assistant play music playlists for me.	AE	−0.51	0.62	−0.49	0.62
27	I use my voice assistant to set my timer (e.g., when cooking or exercising).	AE	−0.57	0.64	−0.18	0.54
28	I allow my voice assistant to access other apps (e.g., WhatsApp, Spotify) and my web activities.	T	−0.67	0.66	0.11	0.47
29	I accept my voice assistant's suggestions.	C	−0.87	0.70	0.15	0.46
30	If my voice assistant misunderstands me, I take the time to give it new input.	C	−0.98	0.72	−0.54	0.63

Note. δ indicates item difficulties in Studies 1 and 2 expressed in logits. Easier items have lower (or more negative) values, whereas difficult items have higher values. p Is the probability that the average person exhibits the respective behavior. AE refers to the category active engagement, RS means reciprocal strength enhancement, T refers to transparency and C to process and/or outcome control.

Table 3. Regression analysis for predicting the use of more than one voice assistant from collaboration intensity.

	B	*SE*(*B*)	β	*t*	*p*
Constant	1.23	0.18		7.02	< 0.001
Gender	−0.12	0.07	−0.10	−1.68	0.093
Technology Acceptance	0.07	0.04	0.11	1.70	0.091
Collaboration Intensity	0.08	0.04	0.12	2.04	0.042

Note. Gender was coded 0 (*male*) and 1 (*female*). The dependent variable was assessed as the number of voice assistants used.

Table 4. Regression analysis for predicting the frequency of voice assistant use from collaboration intensity.

	B	*SE*(*B*)	β	*t*	*p*
Constant	2.21	0.44		5.00	<0.001
Gender	−0.49	0.17	−0.16	−2.82	0.005
Technology Acceptance	0.31	0.11	0.17	2.88	0.004
Collaboration Intensity	0.58	0.10	0.33	5.95	<0.001

Note. Gender was coded 0 (*male*), 1 (*female*).

5.3.1. Construct validity

We used technology acceptance to assess construct validity. The correlation between the Collaboration Intensity Scale and technology acceptance was $r = .32$, $p < 0.001$ (.39, corrected for attenuation). That is, as expected (*H1*), technology acceptance and collaboration intensity were correlated to a moderate degree.

5.3.2. Predictive validity

To test for predictive validity, we assessed whether the scale was adequate for identifying power users (i.e., users who use more than one voice assistant or use voice assistants very frequently). We expected that, when controlling for technology acceptance and users' gender, the Collaboration Intensity Scale would explain additional variance in the use of more than one voice assistant (incremental validity). Table 3 presents the results. When we controlled for gender and technology acceptance, the overall model was significant, $F(3, 313) = 5.59$, $p < 0.001$, $R^2_{(adj.)} = .04$, and the influence of collaboration Intensity on the number of voice assistants used was also significant, however the effect was small ($\beta = 0.12$, $p = 0.042$).

An additional analysis was performed to predict the frequency of voice assistant use (from 1 = *several times a month* to 5 = *several times a day*). Table 4 presents the regression model. When we controlled for gender and technology acceptance, collaboration intensity emerged as a significant predictor ($p < 0.001$) and accounted for additional variance ($\beta = 0.33$), indicating a moderate positive relationship between collaboration intensity and the frequency of voice assistant use, $F(3, 291)$, $p < 0.001$, $R^2_{(adj.)} = .21$. Thus, the data were consistent with H4. We can thus conclude that the Collaboration Intensity Scale is suitable for identifying voice assistant power users beyond their personal degree of technology acceptance.

5.4. Discussion

Study 1 showed that our newly developed scale can measure human-VA collaboration intensity reliably with 30 items and confirmed the possibility of using the entire pool of 70 items for an adaptive test. Construct validity revealed a

moderate correlation between collaboration intensity and technology acceptance, demonstrating that the two variables are conceptually similar, yet measure different constructs. In terms of predictive and incremental validity, the scale demonstrated its worth in predicting the frequent use of voice assistants among users. The relationship between collaboration intensity and the use of multiple voice assistants was also significant. However, the effect observed was rather small and similar in magnitude to the effect of technology acceptance, suggesting that the use of multiple voice assistants may be more dependent on external factors (e.g., the presence of different systems on different devices) than on user-related aspects. The examination of item difficulties provides further insights into particularly difficult behaviors, such as allowing a VA to make payments, which might require a higher degree of trust. Shopping with a VA is difficult because—as we think—it still is quite inconvenient. Thus, the distribution of item difficulties can inform the design of VA functions to design easy-to-use and trustworthy systems (e.g., Rheu et al., 2021). Easier items include that users take the time to provide new input to the VA when it misunderstands them. This indicates that many users—even those with medium collaboration intensity—show a general willingness to support the system, understand its errors and aim to help it improve.

Despite these promising results, the study's sample was relatively homogeneous in terms of age, gender, and education level, which might limit the generalizability of the findings. Prior research has shown that individual differences, such as age and gender, can influence technology affinity and adoption. For instance, age is weakly negatively related to affinity for technology interaction, along with significant gender differences, indicating that men tend to show a higher affinity than women (Franke et al., 2019). This suggests that younger and/or male individuals may engage more intensively in human-VA collaboration. However, our data showed no significant correlations between age or gender and collaboration intensity. Other research also suggests that a female gender was associated with higher usage intentions of chatbots (Zabel & Otto, 2021), showing that gender effects may not be uniform across different technologies and should be examined more closely.

To address this limitation, Study 2 was conducted with a more diverse sample to ensure the scale's validity across different demographic groups.

6. Study 2

When designing and conducting the study, we ensured compliance with ethical standards (APA, 2017; WMA, 2013). As described above, no ethical approval was required (Deutsche Forschungsgemeinschaft, 2023).

6.1. Sample and procedure

Data were collected between August 24 and September 2 2021 with the help of a research panel (Respondi, https://respondi.de). Again, participants were informed about the study's purpose, the voluntary nature of their participation, and their right to withdraw at any time without providing a reason. They then confirmed their consent to participate in the study. Participants were incentivized by the market research company. The sample was representative of the online German population in terms of gender, age, education level, and state. To exclude participants, we followed the same rules as in Study 1. The resulting sample consisted of $N = 321$ participants, which allows a stable item calibration in Rasch modeling (Linacre, 1994). Their ages ranged from 18 to 65 with a mean age of 40.84 ($SD = 13.37$), and $n = 160$ were women. Regarding the highest level of education attained, the sample included 49 participants with a university degree, 48 with a high school diploma or equivalent, and 92 with a primary or secondary school leaving certificate.

6.2. Measures

If not stated otherwise, all items were assessed using a 5-point scale ranging from 1 *(do not agree at all)* to 5 *(fully agree)*.

6.2.1. Collaboration intensity

We used the 30-item scale obtained from Study 1 to assess collaboration intensity. All items were assessed on 5-point scales to enable participants to make a more nuanced assessment (1 = *never*, 5 = *very often/always*, or 1 = *strongly disagree*, 5 = *strongly agree*, respectively). Consistent response categories across the scale were chosen to increase simplicity for the user. The scores were then dichotomized (cf. Study 1).

6.2.2. Commitment to informational self-determination

The correlation between the Collaboration Intensity Scale and the commitment to informational self-determination served as an indicator of construct validity. Therefore, 20 items from the Informational Self-Determination Scale (Bücker et al., 2022) were used. As described above, the items were dichotomized to perform a Rasch analysis. The person separation reliability was .68, and an example item is "Before installing a new app or program, I check the access rights."

6.2.3. Openness to experience

A German translation of the openness to experience domain from the HEXACO-60 personality inventory (Ashton & Lee, 2009; Moshagen et al., 2014) was used to measure openness, which we hypothesized would be positively related to collaboration intensity. A higher expression of this characteristic reflects a high level of interest in new experiences, a greater willingness to learn something new, and an open-mindedness toward unfamiliar situations (Ashton & Lee, 2007). An example item is "I'm interested in learning about the history and politics of other countries." Internal consistency was measured with Cronbach's α (α = .75).

6.2.4. Extraversion

We used 10 items from the HEXACO-60 personality inventory (Ashton & Lee, 2009) to assess extraversion as a trait

that is related to collaboration intensity. The scale showed good internal consistency ($\alpha = .76$). An example item is "I sometimes can't help worrying about little things."

6.3. Results

With the data from Study 2, we successfully calibrated the Collaboration Intensity Scale based on the Rasch Model. All MS Infit Values were between 0.8 and 1.2 and the EAP reliability was .88. We examined correlations between collaboration intensity and age as well as gender. The correlation between the Collaboration Intensity Scale and gender was small but significant ($r = -.11$, $p = 0.044$), indicating that male participants reported a slightly higher collaboration intensity than female participants. The correlation between collaboration intensity and age was not significant.

6.3.1. Construct validity

We assessed construct validity by computing the correlation between the Collaboration Intensity Scale and commitment to informational self-determination. At $r = -.19$ with $p < 0.001$ ($-.25$ corrected for attenuation), the correlation was weak and thus consistent with H2.

6.3.2. Relationships with personality traits

We further tested whether the personality traits openness and extraversion are suitable for predicting the Collaboration Intensity Scale values in a regression. Table 5 shows the results of the multiple regression analysis. The overall model was significant, $F(2, 318) = 5.31$, $p = 0.005$, and the proportion of variance in collaboration intensity that was explained by openness and extraversion (adjusted for the number of predictors) was $R^2_{(adj.)} = .03$. Extraversion emerged as a significant predictor of collaboration Intensity ($\beta = 0.13$, $p = 0.022$), whereas the path from openness to collaboration Intensity was only marginally significant ($\beta = 0.10$, $p = 0.064$). Therefore, the data were partially consistent with H3.

6.3.3. Short scale

Based on the selection criteria described above, we have created a short 12-item version of the scale that can be applied when time constraints are a concern. With a duration of approximately one and a half minutes, this short scale allows for economic testing, while still demonstrating good reliability: The EAP reliability was .77, and the person parameters of the short scale and the 30 item scale were correlated with $r = .93$ ($p < 0.001$). The item difficulty distribution of the short scale is presented in Appendix B.

Table 5. Regression analysis predicting collaboration intensity from openness and extraversion.

	B	SE(*B*)	*β*	*t*	*p*
Constant	−1.67	0.53		−3.14	0.002
Openness	0.24	0.13	0.10	1.86	0.064
Extraversion	0.31	0.13	0.13	2.31	0.022

Note. $R^2_{(adj.)} = .03$.

6.4. Discussion

The aim of Study 2 was to validate the Collaboration Intensity Scale based on a heterogeneous sample, to assess its relation to personality traits, and to select items for a short scale. The internal validation of the 30-item scale showed a high reliability and item difficulties were distributed similarly to study 1. Based on these results, 12 items were selected for a short scale with a good reliability. Construct validation revealed a negative correlation between our scale and commitment to informational self-determination, which was in line with our expectations. Bücker et al. (2022) found that informational self-determination was negatively associated with the actual disclosure of personal data, which is—to some extent—necessary for the intensive use of voice assistants. This correlation raises questions about how users balance convenience and privacy when engaging with these technologies. Do users with a high collaboration intensity develop strategies to mitigate privacy risks, or do they trust the voice assistant more, and accept potential risks?

We further assessed the relation between the traits of openness and extraversion with collaboration intensity. Openness to experience exhibited a marginally significant relationship with collaboration intensity. Previous studies on technology acceptance and use found a positive relation with openness (e.g., Joshi et al., 2023; Svendsen et al., 2013), whereas studies on the relation between openness and attitude towards AI present mixed findings—some studies found small positive correlations (Bergdahl et al., 2023; Sindermann et al., 2022), while others found no correlation (e.g., Schepman and Rodway, 2023). In terms of VAs, openness could have a greater influence on the initial decision to use a new technology (i.e., the question of whether to try it out at all) than on the actual collaboration with such a system. Extraversion emerged as a significant predictor of collaboration intensity. This personality trait, which is related to stronger social interactions (Argyle & Lu, 1990), thus also seems to shape the desire to interact and collaborate with nonhuman (but voice-based) systems, indicating that the construct of collaboration intensity differs from technology acceptance and use, as it incorporates social aspects of the collaboration.

7. General discussion

The aim of this study was to develop a behavioral scale for measuring human-voice assistant collaboration Intensity. The scale, which was based on IRT, provides a novel approach for understanding individuals' collaboration intensity by assessing collaboration behaviors and can contribute to the existing research in the field of human-computer interaction in the following ways. First, we developed and validated a novel scale that can reliably measure human-voice assistant collaboration intensity in different test settings: (a) a 30-item version, (b) a short scale with 12 items for economic testing, and (c) the possibility to use the entire pool of 70 items for an adaptive test. Our results demonstrate construct validity and predictive validity, and we found a positive relation between extraversion and collaboration intensity, which indicates that collaboration with voice

assistants can be distinguished from technology acceptance and use at the construct level and that the former also contains social aspects. Second, by applying IRT, the scale allows for the precise assessment of individuals based on their intensity to collaborate with VAs and enables item-level analyses to identify areas where collaboration might still be difficult for users. Those findings contribute to the literature on technology acceptance (Venkatesh et al., 2003; Venkatesh & Davis, 2000) by providing nuance to the backend (i.e., behavior) of traditional technology acceptance models that usually just measure use (intention) in a general way. Furthermore, IRT enables the scale to be extended, thus, new developments in the field of VAs can be taken into account. While our scale focuses on voice-based collaboration technology, our scale can be easily adapted or extended to collaborations with other robot types such as embodied social robots (Blaurock et al., 2022) or AI-enabled mobile devices (Epstein, 2015; Goodrich & Schultz, 2007). The current fast paced development of generative AI increases and normalizes easy collaboration of humans and different types of robots (incl. VAs) through natural language. Our items cover behaviors that are easily exhibited with other types of robots. For example, accepting a suggestion (item 29) or asking several times before deciding (item 17) could also apply to human-robot collaboration with healthcare or manufacturing robots (Goodrich & Schultz, 2007). Third, the scale focuses on actual and past collaboration behaviors, thereby allowing for a more objective measurement of the construct in comparison with self-report attitude measures (Otto et al., 2018).

7.1. Practical implications

Applying the Collaboration Intensity Scale has several practical implications. By determining the difficulty of a collaboration behavior, it is possible to predict whether a given individual (e.g., someone with a moderate collaboration Intensity) would likely engage in similar behaviors in the future (i.e., whether they would use comparable technologies or features). This knowledge can inform the design and implementation of VAs, CAs, and other collaborative technologies which are rapidly emerging in the past two years since the launch of ChatGPT in November 2022. By reducing barriers and implementing more features that are easy to use, developers can improve the overall user experience and, in turn, it is likely that users will show more collaborative behaviors and, in so doing, support technology improvement. Additionally, our scale may provide information about potential biases in the design of voice assistants. The analysis of interindividual differences in collaboration intensity could help to identify functions which are less accessible for certain user groups. Collaboration intensity data could also be used to design VAs with levels of collaboration tailored to the users' preferences and behavioral patterns. Human-AI collaboration varies from simple substitution through AI to augmentation—where AI and humans complement one another and strengthen each others' strengths—to fully integrated assemblages—where AI and humans function as a dynamic unit (Dellermann et al., 2019; Maedche et al., 2019). By considering users' collaboration intensity, developers can optimize the collaboration, ensuring that AI-based systems support users in decision-making processes in a way that is aligned with their comfort levels and expectations. Moreover, in the workplace context, identifying employees' collaboration intensity might be helpful for increasing person-job fit, as the Collaboration Intensity Scale can help identify individuals who are likely to be comfortable collaborating on tasks with collaborative technology. Assigning individuals with high collaboration intensity to tasks that require intensive collaboration with a VA (and potentially collaborative technologies in general) can facilitate the collaboration and might even result in higher-quality work outcomes. For example, there are multiple use cases for VA in machine maintenance, allowing mechanics to convey with VAs to analyze the problem while having their hands free to work on the machines (Purplescape, 2022).

Beyond promoting collaboration intensity, it is important to ensure that data protection measures of VAs are transparent and manageable for users. Given that users with a high collaboration intensity may be more likely to share personal data, it raises ethical concerns in terms of increased power of and dependency on such systems, requiring sufficient corporate digital responsibility (CDR) of the technology providing companies (e.g., Wirtz et al., 2023) and regulatory frameworks to protect users' information without restricting their ability to engage with VAs.

7.2. Limitations and future research directions

The study's design presents some limitations that may be addressed by future research. The scale is specifically designed for individuals who already use VAs, thus, it might not capture the full spectrum of collaboration intensity with VAs and collaborative intelligence in general. It would therefore be useful to extend the scale to other collaborative technologies such as chatbots, embodied robots, or AI-supported workplace tools. Given that different AI systems vary in their interactivity, it is important to investigate whether there is a general propensity to collaborate with such systems which results in a higher collaboration intensity. Future research might thus expand the scale to other forms of collaborative intelligence or different contexts such as collaboration with AI tools or collaboration with embodied robots in education). Also, in its current form, the scale might be too difficult. Hence, including items that address collaboration with other CAs (e.g., chatbots) could improve the accessibility and applicability of the scale. As the pervasiveness of AI systems (e.g., VAs) is continuously increasing, it is likely that the scale will become easier in the future (in the sense that more people will agree with the items), as more users collaborate more intensively with conversational agents. To assess this assumption, the scale could be integrated into longitudinal assessments to examine changes in collaboration intensity over time.

A potential limitation is the dichotomization of responses, which reduces granularity in Collaboration Intensity. This trade-off was deliberate, as behavioral self-reports often suffer from inconsistencies in frequency-based responses, thus, dichotomization was used to reduce measurement error (e.g., Kaiser & Lange, 2021). While future research could explore

alternative methods (e.g., a Partial Credit Model), our approach allows a robust assessment of collaboration intensity.

As in this study we only examined validity cross-sectionally, it would be beneficial to examine the predictive validity of the scale in terms of non-self-reported actual behavior (i.e., objectively measured VA use behavior). Even though self-reports of actual behaviors tend to be less biased than attitudinal self-reports (Otto et al., 2018), future studies could complement this approach by incorporating objective data such as log data from voice assistants. Such an exploration would help determine the extent to which the scale can accurately predict users' engagement with voice assistants in real-life situations. Additionally, besides openness and extraversion which were assessed in the study, future research could explore the role of other personality traits that have been found to relate to attitude towards other AI technologies (e.g., Honesty-Humility; Zabel et al., 2025) in driving user engagement with voice assistants. Moreover, further research is needed to explore the consequences of high collaboration intensity. For example, is it related to the performance of the human-VA dyad in co-creative tasks or to the satisfaction of the human agent? And could high collaboration intensity also have negative effects (e.g., with regard to the disclosure of data)? Particularly, future studies could investigate whether users with high collaboration intensity are more susceptible to persuasion through CAs, or to over-reliance on their recommendations. Additionally, future studies should focus on the interplay between the user's and the system's contributions and their relation to an effective collaboration (e.g., Xue et al., 2024).

Finally, both studies were conducted in Germany. While Study 2 included a heterogeneous sample, access and use of voice assistants may systematically vary across different socioeconomic groups, geographic locations, and people with different cultural backgrounds. Previous studies suggest that demographic factors such as age and gender, as well as interindividual differences in technological affinity, can influence technology adoption and interaction patterns (Attig & Franke, 2019; Franke et al., 2019). Our results show a small negative correlation between gender and collaboration intensity in study 2, but not in study 1, and no significant correlations between age and collaboration intensity in both studies. Future research should explore demographic and cross-cultural differences in VA adoption.

8. Conclusion

The Collaboration Intensity Scale enhances our understanding of how individuals interact with collaborative intelligence systems, such as VAs, and the individual differences in their contributions to a collaboration with such systems. Given the widespread use of conversational agents in various contexts and their dependence on user collaboration for continuous improvement, having a reliable measurement approach is helpful for evaluating and enhancing collaboration with these systems. By developing a scale for measuring human-VA collaboration intensity, we contribute a useful tool for understanding individual differences in user-VA collaboration, which may also be used to predict future collaboration intensity.

Acknowledgements

We would like to thank Julia Wölfl, Catharina Asbach, and Laura Seidel for their support in the development of the items and in conducting study 1.

Author contributions

CRediT: **Sarah Zabel**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing; **Marah Blaurock**: Conceptualization, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing; **Marion Büttgen**: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing; **Siegmar Otto**: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of generative AI and AI-assisted technologies

When preparing this manuscript, the authors used ChatGPT (GPT-4o) for translation and editing. Following its use, the authors reviewed and edited the content as necessary and take full responsibility for the final version of this article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical approval

The study was conducted in accordance with the Declaration of Helsinki and participants gave informed consent.

Funding

Sarah Zabel and Siegmar Otto gratefully acknowledge funding provided by the Stiftung Innovation in der Hochschullehre [project number FBM2020-EA-1670-01800].

ORCID

Sarah Zabel http://orcid.org/0000-0001-7056-7377
Marah Blaurock http://orcid.org/0000-0002-1281-754X
Marion Büttgen http://orcid.org/0000-0002-9409-8701
Siegmar Otto http://orcid.org/0000-0001-9645-2586

References

American Psychological Association (APA) (2017). *Ethical principles of psychologists and code of conduct*. APA. https://www.apa.org/ethics/code

Argyle, M., & Lu, L. (1990). The happiness of extraverts. *Personality and Individual Differences*, *11*(10), 1011–1017. https://doi.org/10.1016/0191-8869(90)90128-e

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *11*(2), 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2008). The HEXACO model of personality structure and the importance of the H factor. *Social and Personality Psychology Compass*, *2*(5), 1952–1962. https://doi.org/10.1111/j.1751-9004.2008.00134.x

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Attig, C., & Franke, T. (2019). I track, therefore I walk – Exploring the motivational costs of wearing activity trackers in actual users. *International Journal of Human-Computer Studies*, *127*, 211–224. https://doi.org/10.1016/j.ijhcs.2018.04.007

Attig, C., Wessel, D., & Franke, T. (2017). Assessing personality differences in human-technology interaction: An overview of key self-report scales to predict successful interaction. In C. Stephanidis (Ed.), *Proceedings of the HCI International 2017 – Posters' Extended Abstracts, Part I, CCIS 713* (pp. 19–29). Springer. https://doi.org/10.1007/978-3-319-58750-9_3

Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, *90*, 363–371. https://doi.org/10.1016/j.chb.2018.08.028

Barnett, T., Pearson, A. W., Pearson, R., & Kellermanns, F. W. (2015). Five-factor model personality traits as predictors of perceived and actual usage of technology. *European Journal of Information Systems*, *24*(4), 374–390. https://doi.org/10.1057/ejis.2014.10

Bergdahl, J., Latikka, R., Celuch, M., Savolainen, I., Soares Mantere, E., Savela, N., & Oksanen, A. (2023). Self-determination and attitudes toward artificial intelligence: Cross-national and longitudinal perspectives. *Telematics and Informatics*, *82*, 102013. https://doi.org/10.1016/j.tele.2023.102013

Berkman, M. İ., & Şahin, Ş. (2024). Exploring usability as a formative construct through UMUX: A multi-language approach for Turkish adaptation. *International Journal of Human–Computer Interaction*, *40*(2), 457–481. https://doi.org/10.1080/10447318.2022.2121049

Bitkom (2022). Fact Sheet: Sprachassistenten. https://www.bitkom.org/sites/main/files/2022-10/220914_FS_Digitale_Sprachassistenten.pdf

Blaurock, M., Büttgen, M., & Schepers, J. (2024). Designing Collaborative Intelligence Systems for Employee-AI Service Co-Production. *Journal of Service Research*. Advance online publication. https://doi.org/10.1177/10946705241238751

Blaurock, M., Čaić, M., Okan, M., & Henkel, A. P. (2022). A transdisciplinary review and framework of consumer interactions with embodied social robots: Design, delegate, and deploy. *International Journal of Consumer Studies*, *46*(5), 1877–1899. https://doi.org/10.1111/ijcs.12808

Bokolo Jnr, A. (2024). Examining the use of intelligent conversational voice-assistants for improved mobility behavior of older adults in smart cities. *International Journal of Human–Computer Interaction*, *41*(7), 1–22. https://doi.org/10.1080/10447318.2024.2344145

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Taylor and Francis. https://www.taylorfrancis.com/books/mono/10.4324/9781410614575/applying-rasch-model-trevor-bond-christine-fox https://doi.org/10.4324/9781410614575

Borsci, S., Malizia, A., Schmettow, M., van der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022). The Chatbot Usability Scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. *Personal and Ubiquitous Computing*, *26*(1), 95–119. https://doi.org/10.1007/s00779-021-01582-9

Borsci, S., Schmettow, M., Malizia, A., Chamberlain, A., & van der Velde, F. (2023). A confirmatory factorial analysis of the Chatbot Usability Scale: A multilanguage validation. *Personal and Ubiquitous Computing*, *27*(2), 317–330. https://doi.org/10.1007/s00779-022-01690-0

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, *47*(6), 1191–1205. https://doi.org/10.1037//0022-3514.47.6.1191

Brügger, A., Kaiser, F. G., & Roczen, N. (2011). One for all? Connectedness to nature, inclusion of nature, environmental identity, and implicit association with nature. *European Psychologist*, *16*(4), 324–333. https://doi.org/10.1027/1016-9040/a000032

Bücker, J., Gerdes, R., & Kaiser, F. G. (2022). *Do you bother to change cookie settings? Solving the privacy paradox by integrating behavioural costs*. German Psychological Society. Congress of the German Psychological Society.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3.Auflage, 2011). VS Verlag.

Cai, J., Sun, Y., Niu, C., Qi, W., & Fu, X. (2024). Validity and reliability of the Chinese version of Robot Anxiety Scale in Chinese adults. *International Journal of Human–Computer Interaction*, *40*(13), 3355–3364. https://doi.org/10.1080/10447318.2023.2188535

Cerezo, E., Lacuesta, R., Gallardo, J., & Aguelo, A. (2025). Exploring the use of voice assistants in nursing homes. *International Journal of Human–Computer Interaction*, 1–17. https://doi.org/10.1080/10447318.2025.2470282

Chen, Q., Gong, Y., Lu, Y., & Tang, J. (2022). Classifying and measuring the service quality of AI chatbot in frontline service. *Journal of Business Research*, *145*, 552–568. https://doi.org/10.1016/j.jbusres.2022.02.088

Cheng, Y., & Jiang, H. (2020). Ai-powered mental health chatbots: Examining users' motivations, active communicative action and engagement after mass-shooting disasters. *Journal of Contingencies and Crisis Management*, *28*(3), 339–354. https://doi.org/10.1111/1468-5973.12319

Chui, M., Roberts, R., Yee, L. (2022). Generative AI is here: How tools like ChatGPT could change your business. *McKinsey & Company*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/generative-ai-is-here-how-tools-like-chatgpt-could-change-your-business

DeCoster, J., Iselin, A., M. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, *14*(4), 349–366. https://doi.org/10.1037/a0016956

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Deloitte (2019). Global Mobile Consumer Survey 2019. Ergebnisse für den deutschen Mobilfunkmarkt. *Deloitte*. https://www2.deloitte.com/content/dam/Deloitte/de/Documents/technology-media-telecommunications/Global_Mobile_Consumer_Survey_2019_Studie_Deloitte_Deutschland.pdf

Deutsche Forschungsgemeinschaft (2023). *When do I need a statement by an ethics committee?* https://www.dfg.de/en/research-funding/proposal-funding-process/faq/faq-humanities-and-social-sciences#263154

Dutsinma, F. L. I., Pal, D., Funilkul, S., & Chan, J. H. (2022). A systematic review of voice assistant usability: An ISO 9241-11 approach. *SN Computer Science*, *3*(4), 267. https://doi.org/10.1007/s42979-022-01172-3

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

Epstein, S. L. (2015). Wanted: Collaborative intelligence. *Artificial Intelligence*, *221*, 36–45. https://doi.org/10.1016/j.artint.2014.12.006

Fan, A., Lu, Z., & Mao, Z. (2022). To talk or to touch: Unraveling consumer responses to two types of hotel in-room technology. *International Journal of Hospitality Management*, *101*, 103112. https://doi.org/10.1016/j.ijhm.2021.103112

Feng, S., Halpern, B. M., Kudina, O., & Scharenborg, O. (2024). Towards inclusive automatic speech recognition. *Computer Speech & Language*, *84*, 101567. https://doi.org/10.1016/j.csl.2023.101567

Fernandes, T., & Oliveira, E. (2021). Understanding consumers' acceptance of automated technologies in service encounters: Drivers of digital voice assistants adoption. *Journal of Business Research*, *122*, 180–191. https://doi.org/10.1016/j.jbusres.2020.08.058

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21*(1), 1095. https://scholar.google.de/citations?user=wtnaj88aaaaj&hl=de&oi=sra

Fong, T., Thorpe, C., & Baur, C. (2003). Collaboration, dialogue, human-robot interaction. In R. A. Jarvis & A. Zelinsky (Eds.), *Springer tracts in advanced robotics. Robotics research* (Vol. 6, pp. 255–266). Springer. https://doi.org/10.1007/3-540-36460-9_17

Forbes Technology Council (2020). 12 Ways voice assistants could boost business productivity. *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2020/01/17/12-ways-voice-assistants-could-boost-business-productivity/?sh=7c482fcce966

Fox, J., & Weisberg, S. (2019). *An {R} companion to applied regression* (3rd ed.). Sage Publications. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: development and validation of the Affinity for Technology Interaction (ATI) scale. *International Journal of Human–Computer Interaction*, *35*(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: A survey. *Foundations and Trends® in Human-Computer Interaction*, *1*(3), 203–275. https://doi.org/10.1561/1100000005

Guo, E. (2020). How we taught Alexa to correct her own defects: Self-learning system uses customers' rephrased requests as implicit error signals. *Amazon Science*. https://www.amazon.science/blog/how-we-taught-alexa-to-correct-her-own-defects

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

Henkel, A. P., Bromuri, S., Iren, D., & Urovi, V. (2020). Half human, half machine – augmenting service employees with AI for interpersonal emotion regulation. *Journal of Service Management*, *31*(2), 247–265. https://doi.org/10.1108/JOSM-05-2019-0160

Hering, M., Neef, N. E., Zabel, S., & Otto, S. (2024). *General Artificial Intelligence Attitude (GAIA) item pool – how to measure attitudes toward an ever-changing object*. Center for Open Science. https://doi.org/10.31235/osf.io/3c8sr

Holland, P. W., & Wainer, H. (Eds.). (2012). *Differential item functioning* (0th ed.). Taylor and Francis. https://doi.org/10.4324/9780203357811

Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An introduction to voice assistants. *Medical Reference Services Quarterly*, *37*(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391

Hwang, G., Lee, J. [., Oh, C. Y., & Lee, J. [. (2019). It sounds like a woman. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1–6). ACM. https://doi.org/10.1145/3290607.3312915

Jiang, T., Sun, Z., Fu, S., & Lv, Y. (2024). Human-AI interaction research agenda: A user-centered perspective. *Data and Information Management*, *8*(4), 100078. https://doi.org/10.1016/j.dim.2024.100078

Joshi, A., Das, S., & Sekar, S. (2023). How big five personality traits affect information and communication technology use: A meta-analysis. *Australasian Journal of Information Systems*, *27*, 1–39. https://doi.org/10.3127/ajis.v27i0.3985

Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's paradigm for attitude research. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *14*(4), 351–367. https://doi.org/10.1177/1088868310366452

Kaiser, F. G., & Lange, F. (2021). Offsetting behavioral costs with personal attitude: Identifying the psychological essence of an environmental attitude measure. *Journal of Environmental Psychology*, *75*, 101619. https://doi.org/10.1016/j.jenvp.2021.101619

Kaiser, F. G., & Wilson, M. (2004). Goal-directed conservation behavior: The specific composition of a general performance. *Personality and Individual Differences*, *36*(7), 1531–1544. https://doi.org/10.1016/j.paid.2003.06.003

Kasilingam, D. L. (2020). Understanding the attitude and intention to use smartphone chatbots for shopping. *Technology in Society*, *62*, 101280. https://doi.org/10.1016/j.techsoc.2020.101280

Klein, A. M., Hinderks, A., Schrepp, M., & Thomaschewski, J. (2020, June). Measuring user experience quality of voice assistants. In *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–4). IEEE.

Kortum, P., & Oswald, F. L. (2018). The impact of personality on the subjective assessment of usability. *International Journal of Human–Computer Interaction*, *34*(2), 177–186. https://doi.org/10.1080/10447318.2017.1336317

Lee, S., Oh, J., & Moon, W. -K. (2023). Adopting voice assistants in online shopping: Examining the role of social presence, performance risk, and machine heuristic. *International Journal of Human–Computer Interaction*, *39*(14), 2978–2992. https://doi.org/10.1080/10447318.2022.2089813

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

Leschanowsky, A., Rech, S., Popp, B., & Bäckström, T. (2024). Evaluating privacy, security, and trust perceptions in conversational AI: A systematic review. *Computers in Human Behavior*, *159*, 108344. https://doi.org/10.1016/j.chb.2024.108344

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, *7*(4), 328. https://www.rasch.org/rmt/rmt74m.htm#:~:text=%E2%80%A0%20Peter%20Kruyen%20(2012)%20%22,with%20sufficient%20certainty.%22%20https%3A//

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Archives of Rasch Measurement*, *16*(2), 878. https://www.rasch.org/rmt/rmt162f.htm

Loyd, B. H., & Gressard, C. (1984). Reliability and factorial validity of computer attitude scales. *Educational and Psychological Measurement*, *44*(2), 501–505. https://doi.org/10.1177/0013164484442033

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, *7*(1), 19.

MacKenzie, S. B., Podsakoff, P., & Podsakoff, N. P., (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293. https://doi.org/10.2307/23044045

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants. *Business & Information Systems Engineering*, *61*(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Mariani, M. M., Hashemi, N., & Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research*, *161*, 113838. https://doi.org/10.1016/j.jbusres.2023.113838

Mattessich, P. W., & Monsey, B. R. (1992). *Collaboration: What makes it work. A review of research literature on factors influencing successful collaboration*. Amherst H. Wilder Foundation. https://eric.ed.gov/?id=ed390758

McLean, G., Osei-Frimpong, K., & Barhorst, J. (2021). Alexa, do voice assistants influence consumer brand engagement? – Examining the role of AI powered voice assistants in influencing consumer brand engagement. *Journal of Business Research*, *124*, 312–328. https://doi.org/10.1016/j.jbusres.2020.11.045

Merriam-Webster (2023). Interaction. *Merriam-Webster*. Merriam-webster.com

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschsprachigen Version des 60-Item HEXACO Persönlichkeitsinventars. *Diagnostica*, *60*(2), 86–97. https://doi.org/10.1026/0012-1924/a000112

Mylona, I., Aletras, V., Ziakas, N., & Tsinopoulos, I. (2021). Rasch validation of the VF-14 Scale of vision-specific functioning in Greek patients. *International Journal of Environmental Research and Public Health*, *18*(8), 4254. https://doi.org/10.3390/ijerph18084254

Neyer, F. J., Felber, J., & Gebhardt, C. (2012). Development and validation of a brief measure of technology commitment. *Diagnostica, 58*(2), 87–99. https://doi.org/10.1026/0012-1924/a000067

OpenAI (2023). ChatGPT plugins. https://openai.com/blog/chatgpt-plugins

Otto, S., Kröhne, U., & Richter, D. (2018). The dominance of introspective measures and what this implies: The example of environmental attitude. *PloS One*, *13*(2), e0192907. https://doi.org/10.1371/journal.pone.0192907

Park, J., & Woo, S. E. (2022). Who likes artificial intelligence? Personality predictors of attitudes toward artificial intelligence. *The Journal of Psychology*, *156*(1), 68–94. https://doi.org/10.1080/00223980.2021.2012109

Paschen, J., Wilson, M. [., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons*, *63*(3), 403–414. https://doi.org/10.1016/j.bushor.2020.01.003

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). Alexa is my new BFF. In *Proceedings of the 2017 CHI*

Conference Extended Abstracts on Human Factors in Computing Systems. ACM. https://doi.org/10.1145/3027063.3053246
Purplescape (2022). 6 Use *cases of leveraging voice* AI *assistants for manufacturing*. https://www.purplescape.com/6-use-cases-of-leveraging-voice-ai-assistants-for-manufacturing/
PWC (2018). Consumer Intelligence Series: Prepare for the voice revolution. https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/voice-assistants.html
R Core Team (2020). RStudio: Integrated *development* for R. R Foundation for Statistical Computing. https://www.R-project.org/
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/
Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint, arXiv: 1704.04579, 1–21. https://doi.org/10.48550/arXiv.1704.04579
Revelle, W. (2021). psych: Procedures for Personality and Psychological Research. https://CRAN.R-project.org/package=psych
Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction*, *37*(1), 81–96. https://doi.org/10.1080/10447318.2020.1807710
Robitzsch, A., Kiefer, T., Wu, M. (2020). TAM: Test analysis modules. https://scholar.google.de/citations?user=u1xrsjcaaaaj&hl=de&oi=sra
Rohden, S. F., & Espartel, L. B. (2024). Consumer reactions to technology in retail: Choice uncertainty and reduced perceived control in decisions assisted by recommendation agents. *Electronic Commerce Research*, *24*(2), 901–923. https://doi.org/10.1007/s10660-024-09808-7
Rzepka, C., Berger, B., & Hess, T. (2022). Voice assistant vs. chatbot – examining the fit between conversational agents' interaction modalities and information search tasks. *Information Systems Frontiers*, *24*(3), 839–856. https://doi.org/10.1007/s10796-021-10226-5
Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human–Computer Interaction*, *39*(13), 2724–2741. https://doi.org/10.1080/10447318.2022.2085400
Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). Hey Alexa, what's up? In *Proceedings of the 2018 designing interactive systems conference*. ACM. https://doi.org/10.1145/3196709.3196772
Sindermann, C., Yang, H., Elhai, J. D., Yang, S., Quan, L., Li, M., & Montag, C. (2022). Acceptance and Fear of Artificial Intelligence: Associations with personality in a German and a Chinese sample. *Discover Psychology*, *2*(1). 1–12. https://doi.org/10.1007/s44202-022-00020-y
Sowa, K., Przegalinska, A., & Ciechanowski, L. (2021). Cobots in knowledge work: Human–AI collaboration in managerial professions. *Journal of Business Research*, (*125*), 135–142. https://www.sciencedirect.com/science/article/pii/s014829632030792x https://doi.org/10.1016/j.jbusres.2020.11.038
Splendid Research (2019). *Studie: Digitale Sprachassistenten immer beliebter*. https://www.splendid-research.com/de/news/studie-digitale-sprachassistenten-immer-beliebter/
StageZero Technologies (2022). *The importance of data in voice assistant development*. https://stagezero.ai/blog/importance-training-data-voice-assistants-nlp-development/
Statista (2021). *Voice assistant usage in Germany in 2021, by product and age group*. https://www.statista.com/statistics/1274410/voice-assistants-use-by-age-group-germany/
Statista (2024). *Virtual Assistant Technology - statistics & facts*. https://www.statista.com/topics/5572/virtual-assistants/#topicOverview
Svendsen, G. B., Johnsen, J., A. K., Almås-Sørensen, L., & Vittersø, J. (2013). Personality and technology acceptance: The influence of personality factors on the core constructs of the Technology Acceptance Model. *Behaviour & Information Technology*, *32*(4), 323–334. https://doi.org/10.1080/0144929X.2011.553740
Terzopoulos, G., & Satratzemi, M. (2020). Voice assistants and smart speakers in everyday life and in education. *Informatics in Education*, *19*(3), 473–490. https://doi.org/10.15388/infedu.2020.21
van Pinxteren, M. M. E., Pluymaekers, M., Lemmink, J., & Krispin, A. (2023). Effects of communication style on relational outcomes in interactions between customers and embodied conversational agents. *Psychology & Marketing*, *40*(5), 938–953. https://doi.org/10.1002/mar.21792
Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540
Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926
Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation: R package version 1.0.7*. https://CRAN.R-project.org/package=dplyr
Wienrich, C., & Carolus, A. (2021). Development of an instrument to measure conceptualizations and competencies about conversational agents on the example of smart speakers. *Frontiers in Computer Science*, *3*, 685277. https://doi.org/10.3389/fcomp.2021.685277
Wilson, H. J., Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*. https://hometownhealthonline.com/wp-content/uploads/2019/02/ai2-r1804j-pdf-eng.pdf
Wirtz, J., Kunz, W. H., Hartley, N., & Tarbit, J. (2023). Corporate digital responsibility in service firms and their ecosystems. *Journal of Service Research*, *26*(2), 173–190. https://doi.org/10.1177/10946705221130467
World Medical Association (WMA) (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053
Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, (*8*), 3. https://www.rasch.org/rmt/rmt83b.htm
Xue, J., Niu, Y., Liang, X., & Yin, S. (2024). Unraveling the effects of voice assistant interactions on digital engagement: The moderating role of adult playfulness. *International Journal of Human–Computer Interaction*, *40*(17), 4934–4955. https://doi.org/10.1080/10447318.2023.2227834
Zabel, S., Neef, N. E., Yang, X., Heinze, C. F., & Otto, S. (in press). Assessing attitude towards collaboration with robots in the workplace – One element of optimizing robot engagement. *Communications in Computer and Information Science (CCIS)*. Springer.
Zabel, S., & Otto, S. (2021). Bias in, bias out – The similarity-attraction effect between chatbot designers and users. In M. Kurosu (Ed.), *Human-Computer Interaction. Design and User Experience Case Studies. HCII 2021. Lecture Notes in Computer Science* (Vol. 12764). Springer.
Zabel, S., & Otto, S. (2024). *SustAInable: how values in the form of individual motivation shape algorithms' outcomes. An example promoting ecological and social sustainability* [Paper presentation]. Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI. https://doi.org/10.1145/3613904.3642404
Zabel, S., Pensini, P., & Otto, S. (2025). Unveiling the role of honesty-humility in shaping attitudes towards artificial intelligence. *Personality and Individual Differences*, (*238*), 113072. https://doi.org/10.1016/j.paid.2025.113072
Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10. https://CRAN.R-project.org/doc/Rnews
Zwakman, D. S., Pal, D., Triyason, T., & Arpnikanondt, C. (2020). Voice Usability Scale: Measuring the user experience with voice assistants. In *2020 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)* (pp. 308–311). IEEE. https://doi.org/10.1109/iSES50453.2020.00074

About the authors

Sarah Zabel is a psychologist and an associated researcher at the Department of Sustainable Development and Change at the University of Hohenheim in Germany. Her research explores how humans shape AI and how AI, in turn, influences human behavior, with a special emphasis on social and ecological sustainability.

Marah Blaurock is a lecturer at the Ingolstadt School of Management, Catholic University of Eichstaett-Ingolstadt, Germany. Her research focuses on the design and human perceptions of AI in service co-production and the interplay of digitalization and sustainability. She has been published in leading journals such as JSR, P&M, and IMM.

Marion Büttgen is a professor in Corporate Management at the University of Hohenheim, Germany. She has published articles in leading journals such as the *Journal of Product Innovation Management, Human Resource Management, and Journal of Service Research*. Her research interests include digital transformation, leadership, and technologies in service management.

Siegmar Otto is a Professor at the University of Hohenheim. With his psychological background, he researches sustainable digital transformation. His empirical research on human-computer interaction aims to uncover biases in digital systems, but also to find ways to utilize digital change to enable sustainable development in consumer and organizational contexts.

Appendix A

Table A1. Item difficulties and MS infit for the full item Pool (70 items).

	Category	δ	*MS infit*
I have my voice assistant pay for me.	active engagement	5.12	0.93
I call a taxi with the help of my voice assistant.	active engagement	4.01	0.96
I use my voice assistant to get food delivered.	active engagement	3.71	0.89
I book a hotel or flight with the help of my voice assistant.	active engagement	3.45	0.91
I bought paid apps for my voice assistant.	reciprocal strength enhancement	3.30	0.92
I order products over the Internet directly through my voice assistant.	active engagement	2.34	0.88
I schedule appointments for my family with my voice assistant.	reciprocal strength enhancement	2.33	0.90
I renamed my voice assistant.	process and/or outcome control	2.23	1.01
I use my voice assistant to engage in conversations in a foreign language.	reciprocal strength enhancement	2.01	0.92
I use my voice assistant to find a photo from a specific day or place more quickly.	reciprocal strength enhancement	1.95	0.91
I use my voice assistant to adjust the lights in my vehicle.	active engagement	1.90	0.91
I use my voice assistant to suggest leisure activities.	active engagement	1.80	0.90
By using my voice assistant, I forget how to perform the tasks myself.	reciprocal strength enhancement	1.70	1.03
I use my voice assistant for work.	reciprocal strength enhancement	1.41	1.02
I have instructed my voice assistant to address me by my nickname.	transparency	1.33	1.10
I use my voice assistant to control the heated seats in my car.	active engagement	1.15	0.98
I use my voice assistant while watching a series, film, etc. (e.g., start, stop, fast-forward, and rewind).	control	1.10	0.86
I use my voice assistant to control the air conditioning in my vehicle.	active engagement	0.98	1.05
I have my voice assistant solve computing tasks.	active engagement	0.96	0.92
I have my voice assistant suggest gas stations/charging stations to me.	active engagement	0.93	0.95
I use my voice assistant to suggest restaurant alternatives nearby.	active engagement	0.90	0.91
I dictate messages to my voice assistant and allow it to send them.	process and/or outcome control	0.85	1.00
I dictate messages to my voice assistant and send them myself afterwards.	process and/or outcome control	0.77	1.05
I use my voice assistant to create my shopping list.	reciprocal strength enhancement	0.72	0.82
I have my voice assistant read out my messages (e.g., WhatsApp, SMS, e-mails).	active engagement	0.69	1.06
Before I ask anyone else, I ask my voice assistant.	active engagement	0.63	0.95
I ask my voice assistant several times before making a decision.	process and/or outcome control	0.53	1.08
I check the data recorded by my voice assistant.	transparency	0.38	1.26
I use my voice assistant in public.	active engagement	0.34	0.95
I have my voice assistant check the weather at my destination while I'm in the car.	active engagement	0.25	1.00
I have my voice assistant add appointments to my calendar.	reciprocal strength enhancement	0.15	0.80
I ask my environment to be quiet when I talk to my voice assistant.	none	0.12	1.19
I have my voice assistant translate words for me so I can learn.	reciprocal strength enhancement	0.08	0.93
I operate my smart home applications via my voice assistant (e.g., shutters, lights, heating).	reciprocal strength enhancement	0.08	0.97
The way I treat my voice assistant is similar to the way I treat a human counterpart (e.g., talk, interact).	none	0.01	1.08
I set up a personal voice profile so my voice assistant recognizes who is talking to it.	process and/or outcome control, transparency	0.00	1.03
I say only the most necessary things to my voice assistant.	active engagement	−0.03	0.86
I ask my voice assistant for the latest news.	active engagement	−0.12	0.84
I activate a radio channel via my voice assistant.	active engagement	−0.41	0.96
I discontinue the application if, in my opinion, my goal is not achieved.	process and/or outcome control	−0.52	1.29
I make the final decision when my voice assistant makes a suggestion.	process and/or outcome control	−0.58	1.45
I tell my voice assistant to add something to my to-do list or create a reminder.	active engagement	−0.59	0.86
I am willing to let my voice assistant access even more of my data than what is currently possible in order to use functions that are not yet available.	transparency	−0.59	1.02
If my voice assistant does not understand my instructions directly, I do it myself.	process and/or outcome control	−0.59	1.15
I allow my voice assistant to access all my personal data.	transparency	−0.59	0.98
I use my voice assistant to set my alarm.	active engagement	−0.68	0.90
I allow my voice assistant to analyze my speech so the system can evolve.	reciprocal strength enhancement, transparency	−0.71	1.01
I ask my voice assistant when I don't know something and thus learn new things.	reciprocal strength enhancement	−0.75	0.96
If I want to research something, I use my voice assistant to do so.	active engagement	−0.76	0.96
When I hear a song I like, I use my voice assistant to find out the name of the song.	reciprocal strength enhancement	−0.84	0.95
I use my voice assistant for navigation.	reciprocal strength enhancement	−0.91	1.04

(*continued*)

Table A1. Continued.

	Category	δ	*MS infit*
When I'm driving a car, I use my voice assistant to call someone.	active engagement	−0.92	1.04
I keep my voice assistant permanently turned on.	transparency	−1.03	0.93
I grant my voice assistant access to my contacts.	transparency	−1.35	1.07
I allow my voice assistant to access my location.	transparency	−1.46	0.86
I ask my voice assistant about the weather.	active engagement	−1.52	0.81
I have my voice assistant play music playlists for me.	active engagement	−1.55	0.99
I use my voice assistant to set my timer (e.g., when cooking or exercising).	active engagement	−1.61	0.91
I allow my voice assistant to access other apps (e.g., WhatsApp, Spotify) and my web activities.	transparency	−1.71	1.02
I turn off my voice assistant after every use.	transparency	−1.83	0.93
I accept my voice assistant's suggestions.	process and/or outcome control	−1.91	1.04
If my voice assistant misunderstands me, I take the time to give it new input.	reciprocal strength enhancement	−2.01	1.06
I am satisfied with my voice assistant and would like to use it in the long term.	none	−2.06	0.76
Sometimes when my voice assistant gives me a suggestion, I make a different decision.	process and/or outcome control	−2.26	1.17
I activate my voice assistant manually.	active engagement	−2.27	1.17
I activate my voice assistant by voice command.	active engagement	−2.41	0.82
I would recommend the voice assistant to my acquaintances.	None	−2.47	0.76
I question my voice assistant's answer and search again myself.	process and/or outcome control	−2.70	1.20
I find it easy to operate my voice assistant.	none	−4.63	1.05
I give false information to my voice assistant to protect my data/privacy.	transparency	−5.83	0.80

Note. Items in italics are reversed. Some items fit into more than one category. In this case, the items were assigned to the category to which they fit best.

Appendix B

Table B1. Distribution of item difficulty and MS infit values for the short scale in study 2.

	Category	δ	*MS Infit*
I have my voice assistant pay for me.	AE	3.92	0.58
I call a taxi with the help of my voice assistant.	AE	3.11	0.65
I use my voice assistant while watching a series, film, etc. (e.g., start, stop, fast-forward, and rewind).	C	2.00	0.79
I use my voice assistant to suggest restaurant alternatives nearby.	AE	1.42	0.91
I dictate messages to my voice assistant and allow it to send them.	C	1.41	0.92
I allow my voice assistant to access all my personal data.	T	1.16	0.92
I ask my voice assistant when I don't know something and thus learn new things.	RS	−0.72	0.98
I grant my voice assistant access to my contacts.	T	0.47	0.90
I use my voice assistant to set my timer (e.g., when cooking or exercising).	AE	−0.15	0.99
I allow my voice assistant to access other apps (e.g., WhatsApp, Spotify) and my web activities.	T	0.13	0.94
I accept my voice assistant's suggestions.	C	0.17	0.87
If my voice assistant misunderstands me, I take the time to give it new input.	RS	−0.51	0.96

Note: AE refers to the category active engagement, RS means reciprocal strength enhancement, T refers to transparency and C to process and/or outcome control.