Development of an Instrument to Measure Conceptualizations and Competencies About Conversational Agents on the Example of Smart Speakers

Carolin Wienrich[1]* and Astrid Carolus[2]

[1]Institute Human-Computer-Media, Human-Technique-Systems, Julius-Maximilians-University, Wuerzburg, Germany, [2]Institute Human-Computer-Media, Media Psychology, Julius-Maximilians-University, Wuerzburg, Germany

The concept of digital literacy has been introduced as a new cultural technique, which is regarded as essential for successful participation in a (future) digitized world. Regarding the increasing importance of AI, literacy concepts need to be extended to account for AI-related specifics. The easy handling of the systems results in increased usage, contrasting limited conceptualizations (e.g., imagination of future importance) and competencies (e.g., knowledge about functional principles). In reference to voice-based conversational agents as a concrete application of AI, the present paper aims for the development of a measurement to assess the conceptualizations and competencies about conversational agents. In a first step, a theoretical framework of "AI literacy" is transferred to the context of conversational agent literacy. Second, the "conversational agent literacy scale" (short CALS) is developed, constituting the first attempt to measure interindividual differences in the "(il) literate" usage of conversational agents. 29 items were derived, of which 170 participants answered. An explanatory factor analysis identified five factors leading to five subscales to assess CAL: storage and transfer of the smart speaker's data input; smart speaker's functional principles; smart speaker's intelligent functions, learning abilities; smart speaker's reach and potential; smart speaker's technological (surrounding) infrastructure. Preliminary insights into construct validity and reliability of CALS showed satisfying results. Third, using the newly developed instrument, a student sample's CAL was assessed, revealing intermediated values. Remarkably, owning a smart speaker did not lead to higher CAL scores, confirming our basic assumption that usage of systems does not guarantee enlightened conceptualizations and competencies. In sum, the paper contributes to the first insights into the operationalization and understanding of CAL as a specific subdomain of AI-related competencies.

Keywords: artificial intelligence literacy, artificial intelligence education, voice-based artificial intelligence, conversational agents, measurement

INTRODUCTION

Digitalization offers new opportunities across various aspects of our lives—in work-related and private environments. New technologies are increasingly interactive revealing multifarious potentials on both an individual and a societal level. Digital voice assistant systems, for example, have grown in popularity over the last few years (Hernandez 2021; Meticulous 2021), offering an intuitive way of use: Simply by talking to the device, the user can operate it. Although today's usage scenarios are still limited and voice-based assistants in private households are rather used as remote controls (e.g., to play music or turn on the lights) or for web searches (e.g., for the weather forecast), future usage scenarios suggest that voice-based systems could be omnipresent and ubiquitous in our future lives. Multiple new and more complex use cases will result in more complex interactions involving heterogeneous user groups. For example, both children and older people could benefit from the ease of use in a private environment, at school, or in a nursing or health care context (e.g., ePharmaINSIDER, 2018; The Medical Futurist, 2020). Moreover, in a work-related environment, differently qualified employees could interact with these systems, providing knowledge and support to solve specific problems (Baumeister et al., 2019). Application areas of voice-based assistants are further discussed in (Kraus et al., 2020).

Across all the (future) scenarios, ease of use is one of the most promising features of a voice-based system. The increasingly intuitive usage results in decreasing requirements regarding the users' technical knowledge or capacities. In contrast, the complexity of the technological systems and their engineering increases. Attributable to the penetration of easy-to-use voice-based systems, the gap between usage and knowledge increases and gains importance. In turn, unknowing and illiterate users tend to fundamental misconceptions of the technology, e.g., regarding functional principles, expectations, beliefs, and attitudes towards the technological system. Misconceptions or limited knowledge about digital technologies constrain the effective, purposeful, and sovereign use of technology skills (e.g., Chetty et al., 2018). The user remains a somewhat naive consumer of easy-to-use applications, who tends to interact with the systems mindlessly, blindly trusting their device, unthoughtfully sharing private data, or expecting human-like reactions from the device. The latter might be particularly relevant for voice-based systems directly interacting with users, such as conversational agents (short CA). Referring to a societal level, limited competencies and misconceptions contribute to the biased public debate, which focuses either on the risks or glorifies the use of digital technology (Zhang and Dafoe, 2019; Kelley et al., 2019). As a result, users are far from becoming informed and critical operators who understand the opportunities digitization offers in general, and conversational agents in particular, who know how and when to use them, or who know when to refrain from usage

(e.g., Fast and Horvitz, 2017). The *Eliza-Effect* and the *Tale-Spin Effect* are two prominent examples of misconceptions (Wardrip-Fruin, 2001). When a system uses simple functions that produce effects appearing complex, i.e., *Eliza-Effect*, users might overestimate the capabilities of the system. For example, when a speech-based recommender system gives an advice the user will follow this advice without further verification as because he/she trusts in the correctness of the device. In contrast, when a system uses complex functions that produce effects appearing less complex, i.e., *Tale-Spin-Effect*, users might underestimate the capabilities of the system. Such a recommender system might elicit less credibility and users would disregard (correct) advices from the system.

One approach to address this gap is to develop more self-explanatory systems to provide services for which the user needs no prior knowledge. For a detailed discussion of explainable AI systems refer to (Doran, Schulz, and Besold 2018; Goebel et al., 2018). Another approach focuses on the detection of users' misconceptions or limited competencies to learn about the users and to derive design user-centered learning or training programs. Due to the omnipresence and increasing penetration of conversational agents, one key factor of successful digitalization is yielding users with appropriate conceptualizations about conversational agents and competencies to operate them. Consequently, we decided to focus on user's conceptions and competencies. However, which particular conceptualization and competencies are relevant for a "literate" usage of conversational agents? How can such conceptualization and competencies be measured? To answer these questions, in a first step, a conceptualization of "AI literacy" is transferred to the context of conversational agent literacy. The present study specified conceptualization and competencies recently reported as relevant for developing "AI literacy" in general (Long and Magerko, 2020), for one voice-based conversational agent proxy, i.e., smart speakers. The subdomain is defined as the "conversational agent literacy" (short CAL). Then two methodological parts follow. In the first part, the "conversational agent literacy scale" (short CALS) is developed, constituting a first attempt to measure interindividual differences in the "(il)literate" usage of conversational agents. With our focus on smart speakers, we derived 29 items. 170 participants answered these items. An explanatory factor analysis identified five factors leading to five subscales to assess CAL. Subscales and items were analyzed regarding reliability and student's CAL. In the second part, insights into construct validity and impacts of interindividual characteristics have been tested with a subsample of 64 participants. Thus the present study contributes to a first understanding of CAL as a specific subdomain of AI-related conceptualizations and competencies, which allows a sovereign use of conversational technology to unfold the full potential of digitization (e.g., Burrell, 2016; Fast and Horvitz, 2017; Long and Magerko, 2020). Long and Magerko (2020) offered a collection of competencies that are important for AIL. At least to the best knowledge of the authors, the present paper is the first approach to developing an operationalization of this collection.

RELATED WORK

Conversational Agents

A conversational agent is a computer system, which emulates a conversation with a human user (McTear et al., 2016). The dialogue system manages the recognition of speech input, the analysis, the processing, the output, and the rendering on the basis of AI-related methods such as natural language processing, natural language understanding, and natural language generation (Klüwer, 2011; McTear et al., 2016). CAs employ one or more input and output modalities such as text (i.e., chat agents), speech (voice agents), graphics, haptics, or gestures. Various synonyms such as conversational AI, conversational interfaces, dialogue systems, or natural dialogue systems result in conceptional blurring (Berg 2015). In the following, we refer to "conversational agents" and consider the modality of speech. To be more specific, we focused on smart speakers. Smart speakers allow users to activate the device using an intend-word or wake-word such as "Alexa". After the activation, it records what is being said and sends this over the *internet* to the main processing area. The voice recognition service decodes the speech and then sends a response back to the smart speaker. For example, the speech file is sent to Amazon's AVS (Alexa Voice Services) in the cloud for the Amazon system. Amazon published the underlying speech recognition and natural language processing technology with the service of Amazon Lex. Please refer to that service for more technical details. We refer to conceptualizations and competencies essential for the interaction with and understanding of voice-based conversational agents, specifically smart speakers. Since our target group mainly knows the products of Amazon and Google, the study results are closely linked to these devices. However, the basic principles of the approach presented could also be transferred to the modality of text-based systems, as large parts of the underlying operations are similar for both systems.

Media-Related Competencies: From Digital Literacy to Conversation Agent Literacy

In our modern information society, knowledge about digitization processes and digital technologies becomes increasingly relevant. For about a decade, the responsible and reflected use of digital media has been discussed as a **new cultural technique**, which exceeds literacy and numeracy e.g., (Belshaw, 2011). However, it is not easy to provide an exact and distinct definition of digital competencies as different authors have introduced various meanings and definitions (Baacke et al., 1999; Groeben and Hurrelmann, 2002; Güneş and Bahçivan, 2018; Janssen et al., 2013). For example, Gallardo-Echenique et al., 2015 identified a wide range of concepts and approaches associated with digital competence in a literature review, i.e., digital literacy, digital competence, eLiteracy, e-skills, eCompetence, computer literacy, and media literacy. Early concepts such as computer literacy primarily referred to the ability to use a text-processing program or to search the WWW for information (Shapiro and Hughes 1996). Information literacy focused on the individual's

more profound cognitive processes of information processing, such as the ability to understand, evaluate, and use information effectively regardless of its multimedia form (Oxbrow 1998). More recently, frameworks of digital competencies neglect operation skills and refer to a broader set of abilities, including technical and non-technical skills e.g., Chetty et al., 2018).

The rise of AI requires a further extension of the concept of literacy. In this sense, Long and Magerko (2020, p. 2) define "artificial intelligence literacy" (AIL) as a 'set of competencies that enables individuals to evaluate AI technologies critically; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace'. Long and Magerko (2020) presented a literature review analyzing 150 studies and reports to derive a conceptual framework of AIL. Their framework involved five themes, each characterized by a set of 17 competencies and 15 design considerations (**Supplementary Table S8**), describing multifaceted aspects of AIL. The identified five main themes or guiding questions are: 1) What is AI?, 2) What can AI do?, 3) How does AI work?, 4) How should AI be used?, and 5) How do people perceive AI? The framework offers a collection of competencies essential for AIL but lacks an operationalization of AIL, allowing a valid and reliable measure of the AIL aspects. Although Long and Magerko (2020, p. 10), themselves argue that "...there is still a need for more empirical research to build a robust and accurate understanding", their descriptive framework constitutes a good starting point for research on AI-related conceptualizations and competencies. Since no other frameworks or theoretical concepts of AIL exist, at least to the authors' best knowledge, the AIL-framework serves as a basis for our scale development.

Since AI is employed in many different applications and systems, the present paper focuses on a subgroup of AI-based systems, which have recently become increasingly important in many human-AI interactions: voice-based conversational agents, specifically smart speakers. To the best knowledge of the authors, conceptualizations, and competencies essential for the interaction and understanding of (voice-based) conversational agents have not been considered yet. Thus, we introduce the conversational agent literacy (CAL) as a subdomain of AIL. CAL employs conceptualizations (e.g., perceptions, attitudes, mental models) and competencies (e.g., knowledge, interactions skills, critical reflection skills) about the CA itself and the interaction with the CA. From an HCI perspective, identifying and monitoring CAL is of utmost importance because future usage scenarios suggest (voice-based) conversational agents to be omnipresent and ubiquitous in our lives and involve more heterogeneous user groups (Baumeister et al., 2019). Equipping users with appropriate conceptualizations and competencies with regard to digital technology will allow sovereign interactions with digital technologies (e.g., Burrell, 2016; Fast and Horvitz, 2017; Long and Magerko, 2020) and (voice-based) conversational agents. However, monitoring CAL requires measurements, which provide the individual assessments of CAL and indicate development potentials.

Measuring Media-Related Competencies: From Digital Literacy to Conversation Agent Literacy

In accordance with the various interpretations of digital competencies, the measures are multifaceted and standardized instruments are missing. Jenkins (2006) developed a twelve-factorial tool assessing general handling with media. Among others, it includes play ("When I have a new cell phone or electronic device, I like to try out ... "), performance ("I know what an avatar is."), and *multitasking* ("When I work on my computer, I like to have different applications open in the same time.") (see also Literat, 2014). Porat et al. (2018) used a six-factorial instrument measuring digital literacy competencies. It includes, for example, *photo-visual literacy* ("understanding information presented in an illustration."), *information literacy* ("Identifying incorrect or inaccurate information in a list of internet search results."), and *real-time-thinking literacy* ("Ignoring ads that pop up while looking for information for an assignment."). Other approaches used open-ended questions to assess *information* ("judging its relevance and purpose"), *safety* ("personal and data protection"), and *problem-solving* ("solve conceptual problems through digital means") (e.g., Perdana et al., 2019). Another instrument for measuring digital literacy comes from Ng (2012), who distinguished a *technical dimension* (technical and operational skills for learning with information and communication technology and using it in everyday life), a *cognitive dimension* (ability to think critically about searching, evaluating, and creating digital information) and a *social-emotional dimension* (ability to use the internet responsibly for communication, socializing and learning). In addition, the scale also measures attitudes towards the use of digital technologies. The social-emotional dimension of this approach extends previous measures by considering the interactivity of digital technology. However, the development of Ng (2012) scale is not based on any conceptional or empirical foundation leading to difficulties regarding a valid, reliable, and comparable use of the instrument.

In sum, the approaches aiming to measure digital competencies are rather limited in terms of their conceptual range revealing (for a review: Covello, 2010). Moreover, instruments measuring digital competencies have rarely referred to artificial intelligence literacy or conceptualizations and competencies relevant for the sovereign use of (voice-based) conversational agents. The few studies in the field aim for the assessment of associations and perceptions about AI, referring only to the fifth theme (How do people perceive AI?) of the conceptional AIL-framework mentioned above (e.g., AI in general: Eurobarometer, 2017; Zhang and Dafoe 2019b; Kelley et al., 2019; voice-based conversational agent-specific: Zeng et al., 2017; Lau et al., 2018; Hadan and Patil, 2020). However, an analysis of the items of the instruments revealed that the majority referred to either digitalization in general or specific embodiments such as robots (e.g., Eurobarometer 2017). Consequently, and regarding the latent variable, it remains somewhat unclear what exactly the items measured. Besides, the studies report neither the underlying conceptual framework nor criteria of goodness (e.g., reliability, validity). Alternatively, they used single items instead of validated scales. In sum, the quality of measurements available remained unclear. In the area of AIL-related competencies, and particularly regarding conceptualizations and competencies of voice-based CAs, the development of measures and instruments is still in its very early stages. Until today, the literature review reveals no valid and reliable instrument to assess CAL resulting in a research desideratum the present study focuses on.

In sum, digital competencies are associated with a wide range of concepts and measures. The rise of voice-based conversational agents requires further extend the idea of digital literacy. Knowledge has been shown to be a key factor for using new technologies competently. Assessing the state of knowledge allows the implementation of precisely fitting training and transformation objectives (for an overview: Chetty et al., 2018). In this sense, Long and Magerko, (2020) introduced a broad conceptional framework of literacy in the context of AI. However, research so far has not presented tools or instruments allowing an assessment of interindividual levels of AIL or related subdomains such as CAL. Aiming for a first attempt to close this gap, the present study focuses on.

1) the development of an empirically founded measuring instrument to assess CAL and
2) the investigation of first insights into the validation and the impact of interindividual differences.

Overview of the Present Work

With our focus on smart speakers as a proxy of voice-based CAs, we derived 29 items portraying 16 of the original 17 competencies and four of the original five themes introduced by (Long and Magerko, 2020). 170 participants answered these items. An explanatory factor analysis identified five factors suggesting a different structure than the original framework. Items were assigned to five subscales to assess (voice-based) conversational agent literacy (CAL). Subscales and items were analyzed regarding reliability and student's performances (Part I). Finally, preliminary insights into construct validity and the analyses of interindividual characteristics offered insights into CAL of a subsample of 64 participants (Part II). In sum, the present paper presents the first attempt to quantify and measure CAL as a sub-domain of AIL using an empirically based instrument, which follows the conceptional framework introduced by Long and Magerko (2020).

PART I: CONSTRUCTION OF SCALES, FACTOR ANALYSIS, RELIABILITY ANALYSIS, AND STUDENT'S PERFORMANCES

Methods
Development of the Items
Items were derived from the 16-dimensional framework by Long and Magerko (2020). For each dimension, four researchers

TABLE 1 | Item Pool: Example items of the 16 dimensions of (voice-based) conversational agent literacy on the example of a smart speaker.

1. Dimension what is AI?	
3. Sub-dimension: Interdisciplinarity	
What do you think: Which of the following technical disciplines plays no role in the development of a smart speaker?	A- computer science B- Psychology C - Pharmacy D - mechanical engineering E - sociology
2. Dimension: What can AI do?	
5. Sub-dimension: AI's strength and weakness	
What do you think: Which of the following areas of knowledge can a person answer better than a smart speaker?	A- concrete factual knowledge B- abstract concepts C - past events D - interhuman communication E - reasons for factual facts
3. Dimension: How does AI work?	
12. Learning from data	
What do you think: What role do previous requests play in the smart Speaker's response to a current request?	A- past requests are irrelevant B- past requests enable the smart speaker to better understand current requests C -past requests allow to create a profile about personal preferences D - past requests are not stored and therefore do not play a role for current requests E - past requests only play a role if they were spoken in the same pitch
4. Dimension: How should AI be used?	
16. Sub-dimension: Ethics	
What do you think: Is the data from the interaction with a smart speaker shared with advertising companies?	A- yes, to display more relevant ads based on personal interests B- No, because such data is not meaningful enough C - No, because this is a privacy issue. It is not allowed D - yes, because such data provides a detailed insight into the habits of the user E - yes, but only if the users agree to this in the smart speaker settings

individually 1) comprehended the meaning of the dimensions, and 2) developed and phrased items consisting of a question and answering options. Then, the four researchers collaboratively analyzed their individual initial item pools regarding content validity, redundancy, and comprehensibility of the items. Furthermore, to account for non-expert respondents, the final items and their responses were as simple and as unambiguous as possible. In sum, 29 items were derived, with each dimension being referred to by at least one item. Every item consisted of a question and five answers from which the participant needed to select the correct ones. Each of the five answers could be correct or incorrect and received equal weight when summed to arrive at the final score for each of the 29 items. Across these questions, different numbers of the answers could be correct. Therefore, the aggregate score can be from 0 (no correct answers) to five correct answers per item, with higher scores indicating better-founded

competencies. **Table 1** presents an extract of the items used in the process of questionnaire development.

Subjects
One hundred seventy participants voluntarily engaged in an online survey. They were between 16 and 55 years old (M = 21.06, SD = 5.04), 82.7% were female. Regarding the highest educational qualification, 83.53% have finished secondary school, 4.71% reported vocational training of 4.71% had a bachelor's, and 1.18% a master's degree. 11.12% owned an Amazon Echo device, 4.12% owned a Google Home.

Procedure
Participants were briefly instructed about the general purpose of the study, with the procedure following the ethical guidelines laid out by the German Psychological Association. Participants

answered the 29 questions referring to conversational agent literacy.

Results

The following section presents the factor analysis results and the scale, item analysis, and student's performances on the CAL scores.

Factor Analysis

We conducted a factor analysis of the conceptualizations and competencies questions to gain deeper insights into the factorial structure of (voice-based) conversational agent literacy. Factor analysis is a "multivariate technique for identifying whether the correlations between a set of observed variables stem from their relationship to one or more latent variables in the data, each of which takes the form of a linear model" (Field, 2018, p.1016). There were two possibilities for the analysis: an exploratory factor analysis (EFA) or a confirmatory factor analysis (CFA). The CFA tests specific associations between items and latent variables, which a model or a framework hypothesizes. Without any a priori assumptions, the EFA searches for associations between the items indicating underlying common factors which explain the variation in the data (Field, 2018). Referring to the framework postulated by (Long and Magerko, 2020), a CFA would aim to verify the postulated 4-themes factorial structure or the 16-competencies factorial structure. However, we decided to conduct an EFA as a first step of the empirical analysis of the conceptualization of CAL. An explorative analysis of the factorial solution provides the opportunity to detect factorial solutions deviating from the postulated factorial structure. Nevertheless, if our empirical data reflected the postulated factors, an EFA would reveal a four- or 16-factorial solution.

A principal component analysis was conducted. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO >0.5 (KMO = 0.744; Kaiser and Rice, 1974), and Bartlett's test of sphericity was significant ($\chi^2_{(406)}$ = 1,314.29, $p < 0.000$). On the level of KMO values of the individual items, one item was lower than the acceptable limit of 0.5 (Kaiser and Rice, 1974) With reference to (Bühner, 2011), we decided to keep the items in the analysis. In the results part, we will come back to this item again. Anticipating the results section, the factor analysis will reveal that this item will not be part of the final scales. Consequently, the problem can be neglected. To identify the number of meaningful factors, a parallel analysis was conducted, resulting in ambiguous solutions: on the one hand, a three-factorial solution was indicated, on the other hand solutions with four, five and six factors were also justifiable. The additional analysis of the scree plot (detection of a break between the factors with relatively large eigenvalues and those with smaller eigenvalues) could not clarify this ambiguity. As an ambiguous basis of decision-making is a typical challenge during the process of an EFA, content-related and mathematical argumentation needs to be combined for informed choices (Howard, 2016). Therefore, we decided to discard the three-factorial solutions and maintained the five and six factors. Three factors would have meant an inadequate reduction of the postulated four- or 16-dimensional framework (content-related argument) and a relatively small amount of explained variance (three-factor solution: 33.577%;

five: 44.170; six: 51.125; mathematical argument). Because the resulting factors were hypothesized to be intercorrelated, two promax (oblique) rotations—on five and on six factors—were conducted. Then, the rotated solutions were interpreted following both content criteria (conceptual fitting of the items loading on the factor; conceptual differences between items of different factors) and statistical criteria (factor loadings below 0.2 were excluded, loadings above 0.4/0.5 indicated relevance). Negative loadings were also excluded, as they would indicate reverse item coding, which would not work for conceptualizations and competencies items. The two rotated factor patterns of pattern loadings resulting from the promax rotation of five and six factors are presented in (**Supplementary Table S9**).

When analyzing the two sets of factor loadings, factors 1, two and three were almost identical in both solutions, factors 4 and 5 were roughly the same. The analysis of the sixth factor of the 6-factor solution reveals that most of its items are either absorbed by the third or fourth factor in the 5-factorial solution or are below the threshold of 0.2 factor loadings. Moreover, this factor 6 involves the item with the non-acceptable MSA-value. Consequently, the sixth factor could hardly bring added exploratory value resulting the rejection of the 6-factorial solution and the acceptance of the 5-factorial solution.

To interpret the meaning of the resulting five factors, the items with the highest loadings served as reference values. Then, additional items were added following statistical criteria of factor loadings and content criteria asking for the fit with the overall meaning of the other items. As a result, five subscales with three to six items were derived to reflect CAL. However, we must be careful with the interpretation of the resulting scales, as they result from a very first attempt of scale development in conversational agent literacy. We, therefore, regard the present scale development more as a kind of work-in-progress report, the limitations of which need to be discussed in the discussion section. **Table 2** gives an overview of the preliminary version of the CALS, its subscales, and their items. Additionally, the table presents the original concept by Long and Magerko (2020) as well as the assignment of this study's items derived from it.

Scale and Item Analysis

Table 3 presents the results of scale and item analyses. Cronbach's α (internal consistency) was computed for each subscale (factor) and the total scale. Values were between 0.34 and 0.79, indicating mixed results. With reference to the fact that this paper provides a first insight into the operationalization of CAL (and very first insights into the more general concept of AI literacy), we argue with Nunnally (1978) that "in the early stages of research [...] modest reliability" is acceptable. Thus, we carefully regard the internal consistency of subscales 1 to 3 as acceptable [see also (Siegert et al., 2014), who discuss the challenge of inter-rater reliability in the context of emotion annotation in human-computer interaction]. The fourth and fifth subscale, however, reveal more questionable values (0.34 and 0.35). In sum, we preliminarily maintain these subscales and argue for future optimization (see discussion for more detailed considerations). The average item's *difficulty index p* is 68.84, with mean difficulties ranging from 60.31 (factor 1) to 79.98

TABLE 2 | Empirically derived CALS-subscales and their items compared to the original dimension by Long and Magerko (2020).

CALS-subscales	CAL-items	Original dimension Long and Magerko (2020)	Original assignment of CAL-items
CALS			
1 storage and transfer of the smart speaker's data input CALS	20, 21, 24, 26, 27, 28	1 what is AI?	1–3
2 smart speaker's functional principles CALS	5, 8, 9, 11, 15, 16	2 what can AI do?	4–7
3 smart speaker's intelligent functions, learning abilities CALS	1, 4, 12, 14, 19	3 how does AI work?	8–26
4 smart speaker's reach and potential CALS	2, 6, 22	4 how should AI be used?	27–29
5 smart speaker's technological (surrounding) infrastructure	10, 13, 25		

TABLE 3 | Scale and Item analysis of the CAL.

CAL subscale	Internal consistency: Cronbach's α	Item difficulty: Mean (range)	Item discrimination: Mean (range)
1 storage and transfer of the smart speaker's data input	0.788	60.31 (41.56–66.83)	0.55 (0.28–0.66)
2 smart speaker's functional principles	0.690	79.98 (60.12–89.88)	0.43 (0.25–0.52)
3 smart speaker's intelligent functions, learning abilities	0.623	76.83 (67.76–85.92)	0.36 (0.22–0.60)
4 smart speaker's reach and potential	0.337	62.75 (58.47–66.12)	0.22 (0.15–0.32)
5 smart speaker's technological (surrounding) infrastructure	0.345	64.32 (50.59–71.26)	0.21 (0.20–0.22)
Total scale	0.780	68.84	0.35

Note: Items were not corrected in terms of chance score (chose the correct options by guessing).

TABLE 4 | Intercorrelations of the CALS subscales.

CALS subscales	1	2	3	4	5
1 storage and transfer of the smart speaker's data input	—	—	—	—	—
2 smart speaker's functional principles	0.235**	—	—	—	—
3 smart speaker's intelligent functions, learning abilities	0.290**	0.406**	—	—	—
4 smart speaker's reach and potential	0.210**	0.225**	0.314**	—	—
5 smart speaker's technological (surrounding) infrastructure	0.331**	0.253**	0.225**	0.084	—
Total scale	0.727**	0.678**	0.695**	0.525**	0.515

(factor 2). Corrected item scale correlation indicates item discrimination ranging from 0.22 to 0.541.

Table 4 presents the intercorrelations of the five CAL subscales with each other and with total CAL-score revealing significant correlations across all scales (only exception: subscales four x 5). With the exception of the correlation between the fourth and fifth scales, the correlations indicate medium effects (Cohen 1988). Subscale five does not correlate significantly with the total scale demanding detailed attention in the upcoming steps. Assessment of smart speaker's technological (surrounding) infrastructure (subscale 5) seems to be independent of the conceptualizations and competencies about conversational agents.

Conversational Agent Literacy: Conversational Agent Literacy-Scores

To provide coherent results, which allow interpretation of the participants' level of CAL, we needed to adjust the CAL scores to correct for guessed answers. Each item consisted of a question and five answers resulting in five potentially correct answers and five points to gain. For each answer, participants had to decide whether the answer was right or wrong. Consequently, participants had a 50–50 chance to choose the correct response simply by guessing, resulting in 2.5 points correctly guessed points, on average. To control for an overestimation of the performance, the values of the CALS need to be corrected for chance level. Correct answers have been counted as +1, incorrect answers as −1, resulting in a total range per item from −5 (all options per item were answered incorrectly) to +5 (all options per item were answered correctly). **Table 5** gives an overview of the initial and the corrected CAL-values. As the initial and the corrected CAL scores only differ in absolute values but are comparable regarding relative values, they both indicate conversational agent literacy on an intermediate level—with highest values for subscale one and lowest for subscale 5.

PART II: (INITIAL STEPS OF) VALIDATION AND IMPACT OF INTERINDIVIDUAL CHARACTERISTICS

Method
Subjects

Sixty four participants (71.9% women), between 18 and 55 years old ($M = 23.64$, $SD = 7.13$) participated additionally in the second

TABLE 5 | CAL-scores of the student sample.

CALS	Corrected values mean (SD)	Corrected values range: 5.00 to +5.00	Initial valuesMean (SD)	Initial values range: 0.00 to 5.00
1 storage and transfer of the smart speaker's data input	2.76 (1.31)	−0.67 to +5.00	3.01 (1.20)	0.83 to 5.00
2 smart speaker's functional principles	3.20 (1.17)	−1.33 to +4.67	4.00 (1.06)	0.17 to 5.00
3 smart speaker's intelligent functions, learning abilities	3.14 (1.13)	0.20 to +4.60	3.84 (1.16)	0.40 to 5.00
4 smart speaker's reach and potential	1.68 (1.92)	−2.33 to +5.00	3.14 (1.33)	0.48. to 5.00
5 smart speaker's technological (surrounding) infrastructure	1.85 (1.57)	−2.33 to +5.00	3.22 (1.11)	0.70 to 5.00
Total scale	2.70 (0.85)	+0.13 to +4.30	3.44 (1.17)	0.52 to 5.00

TABLE 6 | The survey structure of PART II.

Survey part	Construct	Sub-dimensions
1. validation	Competency-relation	Affinity for technology, technology commitment
	Attitudes towards smart speakers	Attitude towards smart speakers
2. Impact of interind. Characteristics	Individual characteristics	Self-efficacy in smart speaker usage, general self-efficacy
	Demographical data	Gender, age, ownership of voice assistants

part. 70.3% were students, 3.1% have finished secondary school, 10.9% reported vocational training, 12.5% had a bachelor's, and 3.1% a master's degree. 12.5% owned an Amazon Echo, 1.6% a Google Home.

Procedure

To learn more about underlying characteristics and interindividual differences between participants, we assessed additional variables via standardized questionnaires, including competency-related measures and attitudes towards digital technologies and smart speakers, psychological characteristics, and demographical data (an overview is given in **Table 6**).

Measures

Conversational Agent literacy (CAL) was assessed by the newly developed scale (see Part I), incorporating 23 items (see **Table 1** for example items), which were summarized to five factors. CAL values as an indicator of participants conversational agent literacy were then correlated with the following items.

Competency-related constructs were measured by the German version of the *affinity for technology interaction scale* (short *ATI* scale) by Franke et al. (2019). The 6-point Likert scale includes nine items (e.g., "I like to occupy myself in greater detail with technical systems.") ranging from 1 "completely disagree' to 6 "completely agree'. Items were averaged so that higher values indicated higher levels of technique affinity. The internal consistency of the scale was $\alpha = 0.90$.

In addition, the German version of the *commitment for technology short scale* (short *CT*) by Neyer et al. (2012) was used. Twelve items were answered on a 5-point Likert scale ranging from 1 "completely disagree" to 5 "completely agree". A *total score* and three subscales can be calculated: *technique acceptance*, *technique competence beliefs*, and *technique control beliefs*. Items were averaged with higher values indicating higher levels of technique commitment. The internal consistency of the scale was $\alpha = 0.84$.

Attitudes towards smart speakers (short *ATSS*) were measured by using an adapted version of the *Negative Attitude toward Robots Scale* by Nomura et al. (2006). Across all items, we replaced "robots" with "smart speakers". Since the original items measured negative attitudes, we modified the wording as it was already done in (Wienrich and Latoschik 2021). In correspondence to each emotionally coded word used in the original items, we created a semantic differential scale so that the positive and negative emotions would be linked to high and low values, respectively. For example, we built the semantic differential "nervous/relaxed" instead asking "I would feel nervous operating a smart speaker in front of other people." Analogously to the original scale, the adapted scale consists of 14 items, which are answered on a 5-point Likert scale ranging from 1 "strongly disagree" to 5 "strongly agree" on a semantic differential. The *total score* and three subscales can be calculated: *attitudes toward (a) situations of interaction with smart speakers*, *the social influence of smart speakers*, and 3) *emotions in interaction with smart speakers*. Again, items were averaged, and higher values indicated higher levels of positive attitudes. Internal consistency of the scale was $\alpha = 84$.

Psychological characteristics were measured by an adaptation of the *self-efficacy in the human-robot interaction scale* (short *SE-HRI*) by Pütten and Bock (2018). We replaced "robots" with "smart speakers" (short *SES*). Ten items were answered on a 6-point Likert scale ranging from 1 "strongly disagree" to 6 "strongly agree" with higher values of averaged items indicating higher levels of self-efficacy in smart speaker interaction. The internal consistency of the scale was $\alpha = 0.91$. In addition, general self-efficacy was measured using the *general self-efficacy scale* (short *SEG*) by Schwarzer et al. (1997). Ten items (e.g., "For each problem, I will find a solution.") were answered on a 6-point Likert scale ranging from 1 "completely disagree" to 6 "completely agree". Items were averaged, and higher values indicated higher levels of general self-efficacy. The internal consistency of the scale was $\alpha = 0.82$.

TABLE 7 | Correlations between CALS subscales and technology-related competencies and attitudes.

CALS	ATI	CT: Total	CT: Acceptance	CT: Competence	CT: Control	ATSS: Total	ATSS: Situations	ATSS: Influence	ATSS: Emotions
1 storage and transfer of the smart speaker's data input	0.278[a]	0.254[a]	0.058	0.354[b]	0.132	−0.334[b]	−0.290[a]	−0.333[b]	−0.245
2 smart speaker's functional principles	0.297[a]	0.269[a]	0.200	0.254[a]	0.127	−0.332[b]	−0.222	−0.457[b]	−0.195
3 smart speaker's intelligent functions, learning abilities	0.405[b]	0.176	0.070	0.258[a]	0.046	−0.070	−0.015	−0.161	−0.014
4 smart speaker's reach and potential	0.079	0.044	0.021	0.173	−0.111	0.138	0.170	0.137	0.031
5 smart speaker's technological (surrounding) infrastructure	−0.077	0.032	-0.019	0.148	−0.069	−0.284[a]	−0.247[a]	−0.310[a]	−0.178
Total	0.404[b]	0.305[a]	0.147	0.434[b]	0.068	-0.297[a]	−0.193	−0.397[b]	−0.195

[a]$p < .05$.
[b]$p < .001$; ATI = Affinity for Technology Interaction; CT = Commitment to Technology (subscales: technique acceptance, competence beliefs, control beliefs.); ATSS = Attitudes toward Smart Speaker (subscales: attitudes toward situations of interaction with smart speakers, social influence of smart speakers, emotions in interaction with smart speakers).

Demographic data included gender ("female", "male", "diverse"), age, knowledge of German language, level of education, field of study, and the previous experiences with smart speaker (from 1 "never" to 5 "very often") as well as the ownership of smart speakers (e.g., "Alexa/Amazon Echo" or "Google Home").

Expectations for the Validation of the Questionnaire

Besides the analysis of the reliability, the validation of a newly developed questionnaire is a crucial challenge to meet. To evaluate construct validity, CALS subscales are embedded into a "nomological net" of similar constructs. To gain first insights, we analyze the associations between CALS subscales, scales measuring competency-related constructs and attitudes towards technology (Cronbach and Meehl, 1955). Because the competency-related scales are supposed to measure similar constructs, they shall positively correlate with positive attitudes towards technology (ATI, CT). Being more literate in the area of conversational agents should be associated with increased technology affinity. However, attitudes are not to be equated with competencies, so on the one hand, we do not assume perfect correlations, and on the other hand - depending on the type of attitude - we expect negative and positive correlations.

Results: (Initial Steps of) Validation

Correlations with scales measuring similar constructs establish the first steps towards a nomological net of CAL. **Table 7** reveals correlations between the total score of CAL, the factors, and the chosen variables. As expected, Affinity for Technology (ATI) and commitment to technology (CT) correlated positively with the total CAL scale (CALS Total), indicating overlapping concepts. Furthermore, the medium to large positive correlation between CALS Total and the CT-subscale competence can be interpreted as a first indicator of the validity of CALS. The non-significant correlation of CALS four and CALS five underlines this conclusion as both scales do not explicitly refer to competencies.

Remarkably, the analysis revealed that participants' attitudes toward smart speakers (ATSS) correlated negatively with CALS Total as well as with the subscales CALS 1, CALS 2, and CALS 5.

As with commitment before, CALS four did not correlate significantly with attitudes, the same applies to CALS 3. Thus, while participants' competencies in terms of their knowledge about "storage and transfer of the smart speaker's data input" (CALS 1), its "functional principles" (CALS 2) and its technological (surrounding) infrastructure (CALS 5) are negatively associated with their attitudes towards the devices, the participants' knowledge of the intelligent functions and learning abilities of smart speakers are not associated to their attitudes.

Regarding the subscales of ATSS, attitudes toward situations of interaction with smart speakers and social influence of smart speakers are also negatively correlated with CALS 1, 2, and 5. However, the ATSS-subscale referring to emotions occurring when interacting with smart speakers reveals lower and non-significant correlations.

Results: participants' conversational agent literacy and associated psychological and demographic characteristics.

The final step of our analyses aims for first insights into associations between conversational agent literacy and psychological and demographic characteristics. To analyze the impact of interindividual differences, we compared female and male participants and analyzed the effects of age, smart speaker ownership, smart speaker self-efficacy, and general self-efficacy.

The analysis of gender differences revealed that male participants achieved slightly higher scores on CALS-items (total scale: $M = 3.39$, $SD = 0.41$) than woman ($M = 3.76$, $SD = 0.38$). However, differences across all subscales were small and not significant. Furthermore, our sample's gender ratio was not balanced, so this result should be viewed with caution.

Regarding age, the only significant correlation was found for CALS 5 (*smart speaker's reach and potential*), which correlated significantly positively with age ($r = 0.379$; $p = 0.02$). However, the small age range must be considered when interpreting this result.

Interestingly, the ownership of a smart speaker voice assistant did not result in significantly different CAL scores.

The general self-efficacy did not show any significant correlation with CALS-scales. Similarly, participants' self-efficacy in terms of smart speaker interaction did not correlate with the CAL scales. The only exception is CALS 2 (*smart*

speaker's functional principles), which does correlate significantly positively ($r = 0.341$; $p = 0.006$).

DISCUSSION AND OUTLOOK

The concept of digital literacy has been introduced as a new cultural technique complementing the former predominant focus on media literacy and numeracy (e.g., Belshaw, 2011). With reference to the increasing importance of computer technology and particularly AI, literacy concepts need to be extended to meet new technological and AI-related developments. In this sense, Long and Magerko (2020) proposed a conceptional framework of AI literary (AIL), constituting a good starting point for research on conceptualizations and competencies in the area of digital and AI technologies. Encouraged by their conceptional work, the present paper aims for the development of measurement to empirically assess the postulated conceptualizations and competencies. To avoid the reference point of rather abstract AI applications, this study focuses on voice-based conversational agents. Consequently, the original conceptualization of "AI literacy" is transferred to the context of "conversational agent literacy" (CAL). With the goals of developing a measurement tool for CAL and constituting first steps towards an assessment of interindividual differences in the " (il)literate" usage of conversational agents, part I develops the "conversational agent literacy scale" (short CALS). Part II reveals first insights into the "nomological net" of constructs similar to CAL. Moreover, the first associations between CAL and psychological characteristics are analyzed. The results contribute to the operationalization and quantification of CAL to assess individual conceptualizations and competencies in terms of (voice-based) conversational agents.

Comparison: Empirical Vs Conceptional Framework of Aritifical Intelligence literary

In their conceptional framework of AIL, (Long and Magerko, 2020), postulated five themes (What is AI?; What can AI do?; How does AI work?; How should AI be used?; How do people perceive AI?) and 17 competencies. Four of these AIL-themes and 16 competencies could be transformed into the initial set of 29 items (see **Supplementary Table S10** for the complete list of items; APPENDICES). Although the explorative factor analysis of these 29 items revealed five factors, the factors represent a different content structure than the original AIL-framework has postulated. The overlaps and differences between the original and the empirical factorial structure are discussed below.

Following the newly developed CALS, we begin with **CALS 1** (storage and transfer of the smart speaker's data input; $\alpha = 0.788$). CALS one comprises six items, four of which initially belonging to the dimension "How does AI work?". This question summarizes large parts of the AI-literacy concept and encompasses nine out of 17 competencies, accordingly. CALS one covers three competencies (Learning from Data; Action and Reaction;

Sensors). The two remaining items of CALS one are from the fourth dimension (How should AI be used?), asking for ethical considerations. However, our operationalization of ethics was limited to the passing of data from the interaction with the smart speaker. Consequently, CALS one refers to data input, storage, and transfer of the data input the user generates when interacting with the data. Considering the broad scope of ethics and anticipating this study's discussion, our limitation to data sharing seems to be too limited. Future studies will need to differentiate ethical considerations. **CALS 2** (smart speaker's functional principles; $\alpha = 0.690$) comprises six items, five of them belonging to the "How does AI work"-dimension, again. In terms of the 17 competencies, CALS two involves three: Knowledge Representation, Decision Making and Human Role in AI. The sixth item, however, belongs to the "What can AI do?"-Dimension, asking for areas of knowledge in which humans are superior. In sum, CALS two refers to two main aspects: the resemblance of human information processing (representation of knowledge reasoning, decision making) and the human role in terms of the development of the systems and their possible superiority. **CALS 3** (smart speaker's intelligent functions, learning abilities; $\alpha = 0.623$) includes five items, covering three dimensions and four competencies. One item belongs to the "What is AI"-dimension (competency: Recognizing AI), one to the "What can AI do?"-dimension (AI's Strength and Weakness), and three to the "How does AI work?"-dimension (Machine Learning Steps and Learning From Data). Summarizing this scale, CALS three covers competencies and conceptualizations referring to "intelligent" characteristics of smart speakers, their differentiation from standard speakers, their learning features (machine learning and learning from data), with one item also referring to the possible involvement of humans in data analysis. While CALS 1, 2, and three involve five to six items and show good to acceptable internal consistencies (particularly for this early stage of questionnaire development), **CALS 4** and **CALS 5** are of questionable numerical quality. To avoid false conclusions, we want to emphasize the approach of this study again: the aim was to gain the very first insights into the possibility of making AI-related conceptualizations and competencies measurable. Although we are aware of the shortcomings of the scales, we present the entire process as a first attempt to develop and to use a measuring tool of smart speaker-related literacy. The three items of **CALS 4** (smart speaker's reach; $\alpha = 0.337$) cover three competencies of three original dimensions: Interdisciplinarity of the "What-is-AI"-dimension refers to the multiple disciplines involved in the development of smart speakers; Imagine Future IA of the "What can AI do"-dimension asks for future features of smart speakers; and, Critically Interpreting Data of the "How does AI work?"-dimension enquiring if smart speakers can distinguish their users to process their inputs differently. In sum, CALS four asks for the potential of smart speakers by assessing the multiple facets contributing to their development, their capability in terms of adapting to current users, and their future capabilities. Finally, the three items of **CALS 5** (smart speaker's technological [surrounding] infrastructure; $\alpha = 0.345$) refer to the technological infrastructure, and smart speakers are embedded in. All items

are part of the "How does AI work?"-dimension, covering the competencies Decision Making (dependence on internet connection), Machine Learning Steps (hardware used for data storage) and Sensors (sensor hardware). As a result, CALS five asks for the conceptualization of the technological infrastructure, smart speakers depend on.

To summarize the factorial analysis, referring to the original themes, the empirical structure does not seem to reflect the postulated structure closely. However, on the level of the 16 competencies, one might arrive at a different result: of the 16 competencies (29 items) that were entered into the factorial analysis, 11 competencies (23 items) are included in the final scales. Therefore, one can cautiously conclude that the newly developed CALS reflects the original framework, which has been transferred from AI to smart speaker literacy, quite convincingly. Except for Machine Learning, items of a certain competency are only ever taken up by one of the five CALS-scales. Concerning the very early stage of this process and the cautious interpretation of results, we carefully present the first attempt to operationalize and quantify conversational agent literacy.

First Insights into Construct Validity and Students' Performance in Conversational Agent Literacy Scale

When analyzing the correlations of CALS-scales and instruments measuring technology-related attitudes, results revealed substantial overlaps indicating first signs of a nomological net, the newly developed instrument is embedded: Affinity with Technology was significantly positively associated with CALS 1, 2 and 3, Commitment to Technology (total scale) with CALS one and CALS 2, and Attitudes towards Smart Speakers (total) was significantly negatively correlated with CALS 1, 2, and 5. Knowledge about smart speakers seems to be negatively associated with attitudes about these devices, indicating that a positive view of technology does not guarantee technological competencies—but quite the opposite, perhaps. Correlations of CALS Total confirm the significant correlations of the CALS-subscales. CALS 4, however, did not significantly correlate with any other construct, which could be interpreted as another indicator of its questionable quality. Or, it might point to the different scope of CALS 4. With its focus on the (future) potential and the reach of smart speakers, CALS four might be less close to attitudes towards technology. Future studies should widen the nomological net and incorporate more diverse constructs and variables into the analysis, such as experiences with technology, technological competencies, or psychological variables associated with technology-related competencies (among others: underlying motivations of usage and non-usage, personality traits such as openness to new experiences, or curiosity) (e.g., Jenkins, 2006; Literat, 2014; Porat et al., 2018; Perdana et al., 2019). Moreover, future studies should also consider associations with behavioral indicators. Additionally, the fifth AIL dimension ("How do people perceive AI?") could be considered in future CALS-versions.

Since the present data reveal correlations only, future studies should investigate hypotheses about predictors, moderators, and mediators of CAL.

After the CAL-scores were corrected for guessing, our sample ($N = 170$) revealed an intermediated level of conversational agent literacy. The minor interindividual differences seem to mirror the homogeneity of our predominantly student sample. A more heterogeneous sample would probably reveal more detailed interindividual differences. Remarkably, participants who own a voice assistant did not score higher in CAL, indicating that ownership does not guarantee competencies. Smart speakers are easy and intuitive to use and therefore accessible for broad user groups. However, this easiness might create rather positive attitudes and a deceptive impression of an "innocent" technology discouraging users from education. More informed and more critical operators would understand the opportunities digitization offers and would know how and when to use it or when to refrain from usage (e.g., Fast and Horvitz, 2017; Long and Magerko, 2020). Within the other application areas of more complex AI-related systems such as automatic driving, this gap between easy-to-use and underlying technological complexity might even increase. However, the interpretations of our results are still on a speculative level and call for more empirical data.

Limitations and Future Attempts

Conceptional work on AIL and the development of corresponding measuring instruments are in their early stages. This paper presents the first step towards a reliable and valid measure of CAL to allow very first insights into the more general concept of AI literacy. Future studies in this area should consider the following limitations.

The present paper investigates conceptualizations and competencies about smart speakers as one representative of voice-based conversational agents. Although the basic technological operation and interaction principles are transferable to further variations such as text-based CAs, these preliminary results are limited to smart speakers. Future studies should also involve further (voice-based) CAs or AI-related applications. When presenting the 29 items to our sample, we did not differentiate various devices and applications or a specific context but simply referred to smart speakers. Consequently, we do not know the participants' exact reference points (e.g., specific agents, specific domains such as medicine, commerce, assistance). Future studies should differentiate the devices, applications, and contexts of usage.

Moreover, different methodological approaches can be used to learn about user's conceptualizations and competencies, offering potentials for future studies. For example, participants can be observed when interacting with smart speakers in specific use cases. However, direct observations have limitations. First, researchers create arbitrary user interactions in a laboratory—particularly the usage of conversational agents in controlled lab studies is artificially and covers limited use cases. Second, the results are limited for the specific situation presented to the participants. Third, participants must show up in the lab limiting access for many user groups, which is an even severe issue in a pandemic. The aim of the present approach was a different one. We focus on getting insights about different aspects corresponding with the usage of conversational agents in general.

In the past, digital literacy was researched a lot. However, our analyses showed a lack of research regarding AI literacy—a concept with increasing importance. Other approaches have addressed this gap by making systems more self-explanatory. For a detailed discussion about the explainable AI systems, refer to (Doran et al., 2018; Goebel et al., 2018). In contrast, our approach refers to detecting misconceptions and lacks of competencies to better understand users and design user-centered learning or training programs.

The analysis of scales and items revealed potentials for improvement. Along with the five CALS scales, future studies should develop and test further complementing items. Particularly, the fourth and the fifth subscale indicate questionable internal consistency. Future re-analyses and improvements of CALS four should elaborate the principles of the reach, and the potential smart speakers have and will have in the future more profoundly. Therefore, additional items should be developed and tested. Moreover, our operationalization of ethical aspects was far too limited (see CALS 1) and needs to be substantially expanded.

Finally, the conceptualization of our items resulted in a 50–50 chance to simply guess the correct answer (on average: 2.5 points of five points). Future work should consider different conceptualizations and response formats such as multiple-choice questions instead of correct/incorrect questions. The interpretation of CALS scores offers first insights into different areas of conversational agent literacy and the performance levels in the different areas, which could be precisely addressed by training programs explainable AI approaches. Future studies could aim to analyze the performance levels of certain (user) groups such as children, older adults, technology enthusiasts, or skeptics to derive standard or norm values. In sum, more data need to be collected to improve both the scale itself and the use of the resulting performance levels.

In sum, as the starting point of this study was a specific conceptional framework (Long and Magerko, 2020) we neglected other conceptualizations of digital competencies, which involve additional definitions and domains. Thus, future studies should consider different concepts of digital literacy, AIL, and CAL (e.g., Jenkins, 2006; Literat, 2014; Vuorikari et al., 2016; Porat et al., 2018; Perdana et al., 2019). The approach presented by Ng (2012) might be promising since the social-emotional dimension extended previous measures by considering the interactivity of digital technology. Following the recent recommendations of the G20, future conceptualizations should include a more diverse set of skills of technical but also non-technical competencies (Lyons and Kass-Hanna, 2019).

Contribution

Artificial Intelligence will transform the way we work and live—involving other human beings and machines (e.g., Chetty et al., 2018; Lyons and Kass-Hanna, 2019; Long and Magerko, 2020). A recent concept paper of the "future of work and education for the digital age" think tank of the G20 stated: "Standardized assessment tools are essential to consistently measure digital literacy, identify gaps and track progress towards narrowing them, especially for the most vulnerable

populations" (p.1). Furthermore: "The G20 is well-positioned to lead this process of developing comprehensive definitions, strategies, and assessment tools for measuring digital literacy. These efforts would include the diverse set of skills—technical and non-technical—that are and will be needed in the future" (Lyons and Kass-Hanna, 2019, p. 11). Similar statements of the EU, and other national governments emphasize the aim of the present paper to develop a first attempt of the empirically founded measuring instrument of (voice-based) conversational agents as an increasingly popular representative of an AI-related application. The deductive developmental procedure of the present paper ensures a theoretical embedding of the instrument as the underlying conceptional framework by Long and Magerko (2020) integrates findings of 150 recently published scientific articles and reports on the topic of AIL.

From an HCI perspective, standardized measurements allow us to gain deeper insights into various individual competencies and attributes to monitor the effects of digitization and the effects of the digital divide. The understanding and the conceptualization of the required competencies are presented as the first steps towards the conceptualization of "literate users" compared to the "illiterate users". To conceptualize these different user types, their different levels of technological competencies need to be analyzed and understood. Moreover, to distinguish between differently literate users, their competence levels need to be operationalized to allow standardized measures. As in other scientific areas, which refer to interindividual differences in competencies or attributes such as cognitive capacities, emotional states, or behavioral tendencies, for example, this study argues for the operationalization and quantification of CAL to allow the assessment of individual competencies in terms of (voice-based) conversational agents. Finally, standardized measurements can accompany user-centered evaluations of the rapidly growing numbers of platforms, which address competencies referring to digital, AI-related, or conversational technologies but lack a scientific standard of quality regarding underlying conceptions, measurements, and conclusions. Finally, reliable and valid diagnoses allow the implementation of user-centered training measures to develop users' digital competencies be it CAL or AIL (Chetty et al., 2018).

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants

provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CW: concept and structure, theoretical background, scale development, conduction of the study, interpreting results. AC: concept and structure, theoretical background, scale development, analysis of the study, interpreting results.

REFERENCES

Baacke, D., Kornblum, S., Lauffer, J., Mikos, L., Günter, A., and Thiele, A. (1999). *Handbuch Medien: Medienkompetenz. Modelle Und Projekte*. Bonn, Germany: MEDIENwissenschaft: Rezensionen/Rev. doi:10.17192/ep2000.2.2792

Baumeister, J., Sehne, V., and Wienrich, C. (2019). *A Systematic View on Speech Assistants for Service Technicians*. Berlin, Germany: LWDA. doi:10.1136/bmjspcare-2019-huknc.228

Belshaw, D. A. J. (2011). "*What Is "digital Literacy? A Pragmatic InvestigationTese (Doutorado). Departamento de Educação*. Durham: Universidade de Durham.

Berg, M. M. (2015). "NADIA: A Simplified Approach towards the Development of Natural Dialogue Systems." in Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Passau, Germany, Editors C. Biemann, S. Handschuh, A. Freitas, F. Meziane, and E. Métais (Springer-Verlag), 144–150. doi:10.1007/978-3-319-19581-0_12

Bühner, M. (2011). *Einführung in Die Test-Und Fragebogenkonstruktion*. Munich, German: Pearson Deutschland GmbH.

Burrell, J. (2016). How the Machine 'thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data Soc*. 3 (1), 205395171562251–12. doi:10.1177/2053951715622512

Chetty, K., Liu, Q., Gcora, N., Josie, J., Li, W., and Fang, C. (2018). "Bridging the Digital Divide: Measuring Digital Literacy." *Economics: The Open-Access. Open-Assessment E-Journal* 12 (23), 1–20. doi:10.5018/economics-ejournal.ja.2018-23

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences. Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, N.J: Lawrence Eribaum Associates. doi:10.4324/9780203771587

Covello, S. (2010). *A Review of Digital Literacy Assessment Instruments*. Syracuse, New York: Syracuse University, 1–31.

Cronbach, L. J., and Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychol. Bull*. 52 (4), 281–302. doi:10.1037/h0040957

Doran, D., Schulz, S., and Besold, T. R. (2018). "What Does Explainable AI Really Mean? A New Conceptualization of Perspectives." in CEUR Workshop Proceedings, Vol. 2071. Bari, Italy: CEUR-WS.

Eurobarometer (2017). "*Special Eurobarometer 460: Attitudes towards the Impact of Digitisation and Automation on Daily Life*". Brussels: European Commission.

Fast, E., and Horvitz, E.. 2017. "Long-Term Trends in the Public Perception of Artificial Intelligence." 31st AAAI Conference on Artificial Intelligence, AAAI 2017, San Francisco, California, USA: Association for the Advancement of Artificial Intelligence.

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. 5th ed. London: Sage Publications.

Franke, T., Attig, C., and Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *Int. J. Human-Computer Interaction* 35 (6), 456–467. doi:10.1080/10447318.2018.1456150

Gallardo-EcheniqueEliana, E., Minelli de Oliveira, J., Marqués-Molias, L., Esteve-Mon, F., Wang, Y., et al. (2015). Digital Competence in the Knowledge Society. *MERLOT J. Online Learn. Teach*. 11 (1), 1–16.

Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., et al. (2018). *Explainable AI: The New 42?* Cham, Switzerland: Springer, 295–303. doi:10.1007/978-3-319-9974010.1007/978-3-319-99740-7_21

Groeben, N., and Hurrelmann, B. (2002). *Medienkompetenz: Voraussetzungen, Dimensionen, Funktionen*. Weinheim and Munich: Juventa-Verlag.

Güneş, E., and Bahçivan, E. (2018). "A Mixed Research-Based Model for Pre-service Science Teachers' Digital Literacy: Responses to 'Which Beliefs' and 'How and Why They Interact' Questions. *Comput. Educ*. 118 (December 2017), 96–106. doi:10.1016/j.compedu.2017.11.012

Hadan, H., and Patil, S. (2020). "Understanding Perceptions of Smart Devices," in *Lecture Notes In Computer Science (Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics) 12063 LNCS*. Editors M. Bernhard, A. Bracciali, L. Jean Camp, S. Matsuo, A. Maurushat, P. B. Rønne, et al. (Cham, Switzerland: Springer), 102–121. doi:10.1007/978-3-030-54455-3_8

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

Howard, M. C. (2016). A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve?. *Int. J. Human-Computer Interaction* 32 (1), 51–62. doi:10.1080/10447318.2015.1087664

Janssen, J., Stoyanov, S., Ferrari, A., Punie, Y., Pannekeet, K., and Sloep, P. (2013). Experts' Views on Digital Competence: Commonalities and Differences. *Comput. Educ*. 68, 473–481. doi:10.1016/j.compedu.2013.06.008

Jenkins, H. (2006). "Confronting the Challenges of Participatory Culture: Media Education for the 21st Century," in *An Occasional Paper on Digital Media and Learning*. Editors D. John and T. Catherine (Chicago, Illinois: MacArthur Foundation).

Kaiser, H. F., and Rice, J. (1974). Little Jiffy, Mark IV. *Educ. Psychol. Meas*. 34 (1), 111–117. doi:10.1177/001316447403400115

Kelley, P. G., Yang, Y., Heldreth, C., Moessner, C., Sedley, A., Kramm, A., et al. (2019). "*Happy and Assured that Life Will Be Easy 10years from Now: Perceptions of Artificial Intelligence in 8 Countries*†". New York, NY: ArXiv. Availablat: https://arxiv.org/abs/2001.00081.

Klüwer, T. (2011). From Chatbots to Dialog Systems. *IGI Glob*., 1–22. doi:10.4018/978-1-60960-617-6.ch001

Kraus, M., Ludwig, B., Minker, W., and Wagner, N. (2020). "20 Assistenzsysteme," in *Handbuch Der Künstlichen Intelligenz*. Editors G. Gö, U. Schmid, and T. Braun (Berlin: De Gruyter), 859–906. doi:10.1515/9783110659948-020

Lau, J., Zimmerman, B., and Schaub, F. (2018). Alexa, Are You Listening?. *Proc. ACM Hum.-Comput. Interact*. 2, 1–31. doi:10.1145/3274371

Literat, I. (2014). Measuring New Media Literacies: Towards the Development of a Comprehensive Assessment Tool. *J. Media Literacy Educ*. 6 (1), 15–27.

Long, D., and Magerko, B. (2020). "What Is AI Literacy? Competencies and Design Considerations." in Conference on Human Factors in Computing Systems - Proceedings, Honolulu, HI, USA, 1–16. New York, NY: Association for Computing Machinery. doi:10.1145/3313831.3376727

Lyons, A. C., and Kass-Hanna, J. (2019). *The Future of Word and Educaton for the Digital Age Leaving No One behind: Measuring the Multidimensionality of Digital Literacy in the Age of AI and Other Transformative Technologies*.

McTear, M., Callejas, Z., Griol, D., McTear, M., Callejas, Z., and Griol, D. (2016). *The Conversational Interface*. Basel, Switzerland: Springer. doi:10.1007/978-3-319-32967-3_17

Meticulous (2021). "No Tit Available at:" Meticulous Market Research. Available at: http://www.meticulousresearch.com/. (Accessed March 15, 2021).

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2021.685277/full#supplementary-material

Neyer, F. J., Felber, J., and Gebhardt., C. (2012). Entwicklung Und Validierung Einer Kurzskala Zur Erfassung von Technikbereitschaft. *Diagnostica* 58 (2), 87–99. doi:10.1026/0012-1924/a000067

Ng, W. (2012). Can We Teach Digital Natives Digital Literacy? *Comput. Educ.* 59 (3), 1065–1078. doi:10.1016/j.compedu.2012.04.016

Nomura, T., Suzuki, T., Kanda, T., and Kato, K. (2006). Measurement of Negative Attitudes toward Robots. *Interaction. Stud.* 7 (3), 437–454. doi:10.1075/is.7.3.14nom

Nunnally, J. (1978). *Psychometric Theory*. 2nd ed.. New York: McGraw-Hill. doi:10.2307/1161962

Oxbrow, N. (1998). Information Literacy - the Final Key to an Information Society. *Electron. Libr.* 16, 359–360. doi:10.1108/eb045661

Pütten, A. R-V. D. E. R., and Bock, N.. 2018. "Development and Validation of the Self-Efficacy in Human-Robot-Interaction Scale (SE-HRI)" 7 (3), 1–30. doi:10.1145/3139352

Perdana, R., Riwayani, R., Jumadi, J., and Rosana, D. (2019). Development, Reliability, and Validity of Open-Ended Testto Measure Student's Digital Literacy Skil. *Int. J. Educ. Res. Rev.* 4 (4), 504–516. doi:10.24331/ijere.628309

Porat, E., Blau, I., and Barak, A. (2018). Measuring Digital Literacies: Junior High-School Students' Perceived Competencies versus Actual Performance. *Comput. Educ.* 126 (July), 23–36. doi:10.1016/j.compedu.2018.06.030

Schwarzer, R., Bäßler, J., Kwiatek, P., Schröder, K., and Zhang, J. X. (1997). The Assessment of Optimistic Self-Beliefs: Comparison of the German, Spanish, and Chinese Versions of the General Self-Efficacy Scale. *Appl. Psychol.* 46 (1), 69–88. doi:10.1111/j.1464-0597.1997.tb01096.x

Shapiro, J. J., and Hughes, S. K. (1996). "Information Literacy as a Liberal Art Enlightenment Proposals for a New Curriculum." *Teaching Uncc.Edu* 31 (2), 1–6. Available at: https://teaching.uncc.edu/sites/teaching.uncc.edu/files/media/article-books/InformationLiteracy.pdf (Accessed March 15, 2021)

Siegert, I., Böck, R., and Wendemuth, A. (2014). Inter-Rater Reliability for Emotion Annotation in Human-Computer Interaction: Comparison and Methodological Improvements. *J. Multimodal User Inter.* 8 (1), 17–28. doi:10.1007/s12193-013-0129-9

Vuorikari, R., Punie, Y., Gomez, S. C., and Van Den Brande, G. (2016). *DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: The Conceptual Reference Model*. Luxembourg: Publications Office of the European Union.

Wardrip-Fruin, N. (2001). "Three Play Effects – Eliza, Tale-Spin, and SimCity. *Digital Humanities* 1–2. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.2025&rep=rep1&type=pdf.

Wienrich, C., and Latoschik, M. E. (2021). EXtended Artificial Intelligence: New Prospects of Human-AI Interaction Research. Available at: http://arxiv.org/abs/2103.15004. (Accessed March 15, 2021).

Zeng, E., Mare, S., Roesner, F., Clara, S., Zeng, E., Mare, S., et al. (2017). "End User Security and Privacy Concerns with Smart Homes This Paper Is Included in the Proceedings of the End User Security & Privacy Concerns with Smart Homes." in Thirteenth Symposium on Usable Privacy and Security (SOUPS), no. Soups: 65–80, Santa Clara, CA, USA, Berkeley, CA: USENIX Association.

Zhang, B., and Dafoe, A. (2019). Artificial Intelligence: American Attitudes and Trends. *SSRN J.*, 1–110. doi:10.2139/ssrn.3312874