Please quote as: Schmitt, A.; Zierau, N.; Janson, A.; Leimeister, J. M. (2023). The Role of AI-Based Artifacts' Voice Capabilities for Agency Attribution. Journal of the Association for Information Systems (JAIS), 24(4), 980-1004.

Journal of the Association for Information Systems

Volume 24 | Issue 4 Article 7

2023

The Role of AI-Based Artifacts' Voice Capabilities for Agency Attribution

Anuschka Schmitt
University of St.Gallen, anuschka.schmitt@unisg.ch

Naim Zierau
University of St.Gallen, naim.zierau@unisg.ch

Andreas Janson
University of St.Gallen, andreas.janson@unisg.ch

Jan Marco Leimeister
University of St.Gallen / University of Kassel, leimeister@acm.org

Recommended Citation
Schmitt, Anuschka; Zierau, Naim; Janson, Andreas; and Leimeister, Jan Marco (2023) "The Role of AI-Based Artifacts' Voice Capabilities for Agency Attribution," *Journal of the Association for Information Systems*, 24(4), 980-1004.
DOI: 10.17705/1jais.00827
Available at: https://aisel.aisnet.org/jais/vol24/iss4/7

Journal of the Association for Information Systems (2023) **24**(4), 980-1004
doi: 10.17705/1jais.00827

RESEARCH ARTICLE

ISSN 1536-9323

The Role of AI-Based Artifacts' Voice Capabilities for Agency Attribution

Anuschka Schmitt,[1] Naim Zierau,[2] Andreas Janson,[3] Jan Marco Leimeister[4]

[1]Institute of Information Management, University of St.Gallen, Switzerland, anuschka.schmitt@unisg.ch
[2]Institute of Information Management, University of St.Gallen, Switzerland, naim.zierau@unisg.ch
[3]Institute of Information Management, University of St.Gallen, Switzerland, andreas.janson@unisg.ch
[4]Institute of Information Management, University of St.Gallen, Switzerland /
Research Center for IS Design, University of Kassel, Germany, leimeister@acm.org

Abstract

The pervasiveness and increasing sophistication of artificial intelligence (AI)-based artifacts within private, organizational, and social realms are changing how humans interact with machines. Theorizing about the way that humans perceive AI-based artifacts is, for example, crucial to understanding why and to what extent humans deem these artifacts to be competent for decision-making but has traditionally taken a modality-agnostic view. In this paper, we theorize about a particular case of interaction, namely that of voice-based interaction with AI-based artifacts. We argue that the capabilities and perceived naturalness of such artifacts, fueled by continuous advances in natural language processing, induce users to deem an artifact as able to act autonomously in a goal-oriented manner. We show that there is a positive direct relationship between the voice capabilities of an artifact and users' agency attribution, ultimately obscuring the artifact's true nature and competencies. This relationship is further moderated by the artifact's actual agency, uncertainty, and user characteristics.

Keywords: Voice, User Perception, Agency, Artificial Intelligence, Human Computer Interaction

Robert Wayne Gregory was the accepting senior editor. This research article was submitted on February 15, 2022 and underwent two revisions.

1 Introduction

With the rise of artificial intelligence (AI)-based artifacts and their increasing prevalence in organizational and day-to-day contexts, scholars have theorized about the extent to which human decision-making can be automated and augmented through AI (Berente et al., 2021; Brynjolfsson, 2022; Gregory et al., 2021; Metcalf et al., 2019; Raisch & Krakowski, 2020), i.e., the adequate agency of an artifact (Baird & Maruping, 2021; Jennings et al., 1998; Puranam & Vanneste, 2022). This is underlined by the conceptual understanding of AI, defined as an artifact's performance of "cognitive functions usually associated with human minds" (Nilsson, 1971, qtd. in Raisch & Krakowski, 2020, p. 192).

Increasingly, AI-based artifacts are being embedded in interfaces that interact with the user via natural language dialogue that can be differentiated by modality, i.e., text or voice (Manning & Hirschberg, 2015; Schöbel et al., in press). In particular, the hands- and eyes-free nature of voice-based interaction offers opportunities for industry applications such as surgery machinery, and for certain groups, such as visually impaired individuals (Branham & Roy, 2019). However, the naturalness of the interaction evoked through a humanlike artifact voice and dialogue—coupled with the complexity and opacity of AI—might induce users to make inappropriate inferences about the artifact's capabilities (Berente et al., 2021; Lin et al., 2017; Malle et al., 2020; Natale, 2021). This becomes particularly apparent when considering

voice-based artifacts such as Amazon Alexa taking breathing breaks, adapting its tone of voice to the context of an interaction, and learning from mistakes from previous interactions (Low, 2020; Tarantola, 2020). Alexa's voice capabilities can make the system appear highly capable—and agentic.

When evaluating the thesis of whether AI-based artifacts' possess agency, one can differentiate between the ontological claim that AI-based artifacts are objects of agency (i.e., actual agency) and the empirical claim that people conceive of such artifacts as objects of agency (i.e., agency attribution). While extant literature has addressed the question of whether AI-based artifacts (should) have agency (e.g., Baird & Maruping, 2021; Siddarth et al., 2021), the empirical study of when and why humans might perceive AI-based artifacts as agentic introduces a novel, important perspective to this discussion (Puranam & Vanneste, 2022). The motivation for our current work was to revisit and expand extant notions of human-computer interaction (HCI) on why and under which circumstances users attribute agency—the ability to act in an autonomous, flexible, and situated manner—to artifacts (Gray et al., 2007, Puranam & Vanneste, 2022).

The sociotechnical perspective of HCI can help explain how humans perceive and interact with artifacts (Zhang et al., 2002). For example, Jussupow et al. (2021) found that users unconsciously consign decision agency to artifacts and thereby also rely on incorrect AI advice, ultimately leading to "a broader but less obvious substitution of human agency in crucial decision tasks" (p. 731). In that sense, agency attribution is an important cognitive step that influences users' downstream behavior, as well as the relationship and future interactions between AI-based artifacts and their users. To date, HCI research has studied the way that users perceive (Diederich et al., 2022), rely on (Fügener et al., 2022), and adapt their behavior to AI-based artifacts (Strich et al., 2021) but has focused on AI-based artifacts as a rather generic, modality-agnostic system class. As "AI technologies … occupy a curious place somewhere between humans and inanimate technology in the extent to which they are seen as agentic" (Puranam & Vanneste, 2022, p. 4), studying specific artifact characteristics, such as interaction modality, is decisive in understanding the underlying cognitive processes. In turn, such an understanding can help explain why and what factors induce users to attribute agency to artifacts.

Specifically, we explore the phenomenon of voice-based user interaction with AI-based artifacts as an emerging and essential facet of HCI (Seaborn et al., 2021). In this paper, we study the role of voice capabilities of AI-based artifacts for agency attribution. Our point of departure is the reasoning that a user's perception of an AI-based artifact involves a sociocognitive process and is formed by a multitude of factors. More precisely, we posit that voice resembles natural human interactions more than any other interaction modality and might thereby project capabilities beyond an artifact's actual agency. Taking a modality-agnostic view, prior information systems (IS) and HCI literature cannot readily explain why, how, and when users deem an artifact to be capable and how voice as an interaction modality exacerbates issues of agency attribution. In this research, we thus examine the role of voice to address the following research question:

RQ: Why and under what conditions is agency attributed to an AI-based artifact through interactions via voice?

We develop a model of voice-based user interactions with AI-based artifacts that complements and extends existing theory on HCI (Zhang et al., 2002; Diederich et al., 2022). Extant work on agency attribution posits that an artifact's (social) cues give rise to certain expectations and attributions about the system, such as trust and the intention to use (Puranam & Vanneste, 2022; Zhao & Malle, 2022). Based on the premise that AI-based artifacts differ significantly from other types of information systems in terms of technical performance and the nature of interaction, a fresh look at agency attribution is warranted (Schuetz & Venkatesh, 2020). In this paper, we propose that considering an artifact's "voice capabilities"—that is, the machine-based synthesis of voice embedded in AI-based artifacts—yields not only a more comprehensive understanding of how human users interact with AI-based artifacts but also a unique understanding of when humans are more likely to attribute agency to an artifact.

Addressing our research question is relevant for three key reasons. First, voice as an interaction modality is fundamentally changing HCI. With voice-based interactions more closely resembling human communication and transmitting rich nonverbal cues about the speaker, they afford a distinctive understanding of how AI-based artifacts impact their users. Coupled with its often disembodied and low-definition characteristics, the voice of AI-based artifacts can act as a potentially hazardous and unjustified multiplier of users' reactions (i.e., agency attribution) to these artifacts (Natale, 2021). Second, agency attribution can act as a mediator for important subsequent perceptual and behavioral outcomes (Puranam & Vanneste, 2022). Understanding users' cognitive processes can shed light on the reasons why attributed agency might deviate from a system's actual capabilities and can potentially explain human behaviors, such as the overreliance on AI-based artifacts. This understanding can, in turn, inform system engineers and designers on how to calibrate users' understanding of an artifact's actual capabilities. Third, the introduction of AI-based artifacts into work systems raises the question of how we should design organizations and associated decision-making

processes in the era of human-AI hybrids (Lebovitz et al., 2022; Rai et al., 2019). While voice-based interactions promise to facilitate and shape work activities more effectively, novel issues of privacy and the previously mentioned implications of overreliance play an important role in organizational contexts (Martin & Finnegan, 2020). Studying the consequences of voice-based interactions, both intended and unintended, is of critical improtance. Ultimately, examining how the introduction of voice as an interaction modality shapes human perception, behavior, and interactions with AI-based artifacts is essential to understanding how HCI is increasingly mimicking the breadth of human interaction.

2 Conceptual Background

Before presenting our conceptual model, we first summarize the relevant prior work on voice and interaction with AI-based artifacts.

2.1 Voice as an Interaction Modality of AI-Based Artifacts

Voice is of utmost importance to human life and interaction because of its profound physiological, psychological, cognitive, and behavioral effects. It acts as humans' primary warning mechanism, is considered the most distracting sound, and substantially influences our mood, emotions, and behavior (Treasure, 2020). In contrast to speech, voice "is the way something is said (pitch, timing, loudness)" (Crumpton & Bethel, 2016, p. 271). Despite its seemingly simple output, voice seems to be considerably more intuitive and multifaceted than other modalities. First, voice-based interaction facilitates natural communication through sequential turn-taking and thus more natural interaction pauses. Due to the higher time resolution associated with hearing, as compared to reading, information can be processed much more quickly through voice (Rubin et al., 2000; Villazon, 2021). Second, voice-based interaction is characterized by more colloquial language and syntax, which allows for the effortless expression and reception of information (Dennis et al., 2008). Third, voice encompasses signals beyond the mere provision of information (Rosenthal & Ryan, 2000). Voice can reveal information about personality, identity, and emotions (Sutton et al., 2019). In fact, voice has been defined as "the carrier of speech" (Belin et al., 2004, p. 129) because it becomes distinguishable from other modalities by signaling social cues beyond the linguistic meaning of the information communicated (Redeker, 1984). Combined, these factors make voice a natural and meaningful interaction modality, even for interaction with AI-based artifacts.

Based on natural language processing (NLP), AI-based artifacts can understand, learn, and respond to human language (Gentile et al., 2011; Hirschberg & Manning, 2015). Voice-based interaction is accessible through stand-alone interfaces such as Amazon Alexa and less tangible artifacts such as automotive user interfaces and can be integrated into everyday devices, e.g., desktop interfaces such as Microsoft Cortana (Balasuriya et al., 2018; Craig & Schroeder, 2017). With automatic speech recognition and natural language understanding, which enable machines to understand human users' intent and translate it into textual data, natural language generation and text-to-speech (TTS) technology convey information back to users and provide them with auditory responses (Pearl, 2016). TTS is enabled by relying on human recorded voice and, more recently, increasingly sophisticated synthetic speech generation. Lee and Nass (2004) coined this type of computer-synthesized speech as "doubly-disembodied language"—i.e., artificially generated speech that is not only disembodied from actual human speakers but no longer originates with humans at all.

As presented in Figure 1, dialogue management connects the processes of automatic speech recognition and natural language understanding by determining users' requests and their appropriate reactions, i.e., performing a requested activity or providing the user with an auditory response (Hirschberg & Manning, 2015). Through steadily improving accuracy in automatic speech recognition, dialogue management, and TTS, the ability to recognize and produce normal human conversational behavior has become prevalent in AI-based artifacts. This development is particularly striking when considering the increasing prevalence of large language models (Vaswani et al., 2017). Software development kits such as Python's NLTK provide end-to-end platforms for combining all of these processes (Bird et al., 2009).

Enabled by this technical interaction process, AI-based artifacts can en- and decode auditory cues from the human voice (Hildebrand et al., 2020; Schmitt et al., 2021; Seaborn et al., 2021). Jurafsky and Martin (2023) provide a good starting point for understanding the components of voice and its translation into AI-based artifacts. They propose four layers—amplitude, frequency, temporality, and quality—that can describe the physical properties of a human soundwave. The amplitude of a soundwave describes the loudness or intensity of a voice with a high mean intensity and determines whether the voice will be perceived as extroverted or assertive (Hess et al., 2009).

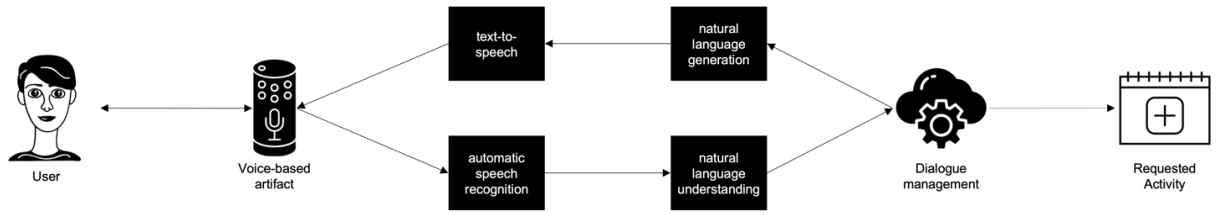

Figure 1. Technical Process of Human Interaction with Voice-Based Artifacts

Frequencies explain how high-pitched a waveform sounds and include formant frequency, which is the lowest harmonic frequency (i.e., pitch) perceivable by humans (Law & Rennie, 2015). Its importance is reflected in the literature as the most observed dimension of all auditory cues (Elkins & Derrick, 2013; Nass & Lee, 2001; Tay et al., 2014). Temporality captures a soundwave in terms of time and includes auditory cues such as the rate of speech (i.e., syllables per second) and pauses (unfilled or filled through nonverbal expressions such as "uh-huh" or "hmm"). Lastly, the quality of a voice is perceivable through the harmonic-to-noise-ratio as the ratio of periodic to nonperiodic components of speech sounds, for instance, which influences the perceived creakiness of a voice. Understanding auditory cues in the human voice enables us to lay out the technical feasibility of creating and modifying voice in AI-based artifacts and identify the differentiating factors of voice employed by AI-based artifacts.

While very few voice-based artifacts fully execute tasks themselves or interact in open domains, the hands- and eyes-free nature of voice offers promising applications for certain user groups, such as children (Aeschlimann et al., 2020), the elderly, (Straßmann et al., 2020), and visually impaired (Abdolrahmani et al., 2018) and disabled people (Masina et al., 2020), as well as for certain use cases such as surgery (Perrakis et al., 2013) and driving scenarios (Truschin et al., 2014), where high cognitive demand is required. At the same time, the adaption of an artifact's voice also leaves room for the manifestation of stereotypes. Previous studies have illustrated how voice genders are attributed to gender-stereotypical tasks, e.g., a male voice for an artifact deployed for a security task (Trovato et al., 2017) or a female voice for an artifact deployed for personal home assistance (Carpenter et al., 2009).

2.2 Human Interaction with AI-Based Artifacts

When discussing an AI-based artifact's "emulation capabilities, i.e., its ability to think humanly" (Berente et al., 2021, p. 1436), the question arises of how capable a human actually perceives an artifact to be. From a sociotechnical perspective, user interaction with AI-based artifacts consists of three key components—the human (i.e., the user of an AI-based artifact), the technology utilized (i.e., an AI-based artifact), and the goal to be achieved (i.e., the task augmented by the AI-based artifact) (Zhang & Li, 2005). In that sense, human interaction with AI-based artifacts hinges on how users perceive AI-based artifacts. Agency attribution is thus a critical cognitive step that influences the relationship and future interactions between users and AI-based artifacts (Puranam & Vanneste, 2022). Interestingly, why we deem an artifact to be capable of executing a task or the extent to which an individual attributes agency to an artifact depends not only on the system capabilities experienced but also on their interplay with contextual factors and the user's beliefs and cognitive heuristics (Jia et al., 2022; Ross et al., 1977).

Considering the technological advancements in NLP, with organizational and individual tasks being increasingly augmented by AI-based artifacts, we can expect that human interaction with AI-based artifacts will become increasingly effortless and natural and thus also voice based. These tendencies warrant a fresh look at human interaction with AI-based artifacts. The predominant auditory interface modality and interaction via spoken language provide a dialogue logic closely aligned with the human senses used in communication, thus resembling human-human interaction more closely than alternative HCI (Cohen et al., 2004; Quesada & Lautenbach, 2017; Zierau et al., 2022). It should be noted that voice-based artifacts not only provide HCI via a different modality but also give rise to novel aspects considered within an interaction, i.e., the richness of information transmitted. The way in which a voice is interpreted significantly depends on users' hearing capabilities (e.g., impaired hearing), their sociocultural associations (e.g., identifying an accent and assigning value to it), and contextual factors (e.g., a familiar environment or a noisy background). Voice-based artifacts are based on the ability to understand and respond to human spoken language and thus also on the ability to create a natural voice. We therefore expect voice-based interactions and the specific auditory cues of AI-based artifacts to give rise to distinct implications for the user. The agency attributed to an artifact would then be a function of the voice capabilities of the artifact.

Research findings building on the computers are social actors (CASA) paradigm and social response theory (Nass et al., 1994; Nass & Moon, 2000) provide a seminal starting point for exploring voice-based interaction with

AI-based artifacts. These theories posit that humans unconsciously apply social heuristics and respond to technological artifacts in a "social manner" comparable to the heuristics applied to humans. Extant studies have shown that voice and its components, such as rate of speed and pitch, directly impact users' perceptions of the artifact or the overall interaction (Chiou et al., 2020; Gálvez et al., 2020). For instance, using a robot in an investment task, Torre et al. (2020), found that a high pitch not only represented the auditory correlate of a smiling voice but also resulted in users placing greater trust in the robot. In an experimental setup of a hotel reservation task, the inclusion of pauses was proven to increase the perceived naturalness of a voice-based artifact and compensated for the lack of logical flow in the interaction (Marge et al., 2010). On the other hand, more natural or humanlike voices do not necessarily always translate into desirable perceptual or behavioral outcomes. While synthetic voices have been associated with less naturalness and social presence than recorded human voices, numerous studies have shown a significant positive effect on trust and preference. Moreover, in learning contexts, synthetic voices have been found to positively affect learning transfer, training efficiency, and engagement (Craig & Schroeder, 2017; Komiak & Benbasat, 2003; Qiu & Benbasat, 2005; Tamagawa et al., 2011). While attributed agency seems to be primarily influenced by the perceived naturalness of an artifact and its voice, other factors seem to influence the agency we ascribe to such artifacts.

In sum, the underlying mechanisms of how voice-based interactions with an AI-based artifact affect the agency users attribute to the artifact influence HCI by modifying and extending the extant interaction behavior, information transmitted, and evoked user perceptions. Accordingly, we develop a conceptual model addressing our research question of why agency attributed to an artifact is driven by the voice capabilities of an AI-based artifact.

3 A Model for Exploring the Role of Voice Capabilities for Agency Attribution when Interacting with AI-Based Artifacts

Figure 2 shows our model for explaining the role of voice capabilities for agency attribution in interactions with AI-based artifacts. This study's explicit focus on voice as a predominant interaction modality for AI-based artifacts is phenomenon driven and motivated by the distinctive implications introduced by voice as an interaction modality. Voice-based interactions are manifested in the positive direct relationship between the voice capabilities of an AI-based artifact and the agency attributed to this artifact by the user interacting with it—a relationship that is moderated by an artifact's actual agency, uncertainty, and the user's characteristics.

The model is based on the following boundary assumptions: First, the "changing expectations and practices of consumers … influence the IT-related activities of workers and managers in organizations" (Gregory et al., 2018, p. 1225) and turn voice-based artifacts into technology that is no longer used solely by consumers in everyday life and for leisure but also by workers in organizational decision-making (Assumption 1). IT consumerization has thus blurred the line between using voice-based artifacts at work or for leisure, with such artifacts being applicable to automating or augmenting *individual-level decision-making* in both contexts (Elshan et al., 2023; Raisch & Krakowski, 2020). Tedious work-related tasks such as booking meeting rooms and starting conference calls are being increasingly automated through voice-based artifacts (Finnegan, 2017). For instance, Alexa for Business can schedule calendar events and transcribe discussion points after meetings (Amazon Web Services, 2022).

Second, while voice is one of several interaction modalities used in HCI, disembodied AI-based artifacts interacting via voice only are increasingly prevalent (Natale, 2021). According to Heider's (1920; 1925) attribution theory and theory of object perception, human perception and the attribution of traits to others are reconstructive processes informed by perceptual cues (Malle & Ickes, 2000). Bringing these two notions together and holding voice constant as the predominant interaction modality, *users reconstruct and hence make sense of AI-based artifacts through perceivable variances in the medium of voice* (Assumption 2)—as compared to the potentially more invariant qualities (i.e., nonobservable, technical qualities) of an AI-based artifact.

Third, in line with the conceptual understanding of AI and its *emulation capabilities*, the degree of AI-based artifacts' competence and ability to interact and execute tasks can be viewed as the degree to which they can mimic human capabilities and skills (Brynjolfsson & Mitchell, 2017; Samuel, 1959). Simultaneously, agency can be viewed as a signal of such ability (Puranam & Vanneste, 2022). This suggests that holding human agency based on age and cognitive ability constant, the agency we would attribute to another human is the highest level of agency attributable to an AI-based artifact (Assumption 3). This is in line with Puranam and Vanneste's (2022) theorizing about agency attribution, in which they compare different types of artifacts and posit that "the regression model is likely to be perceived as having the lowest level of agency and the human, the highest" (p. 11), with conversational and hence voice-based artifacts falling somewhere in between. Drawing on this set of assumptions, we explain our conceptual model in the following sections (also see Figure 2).

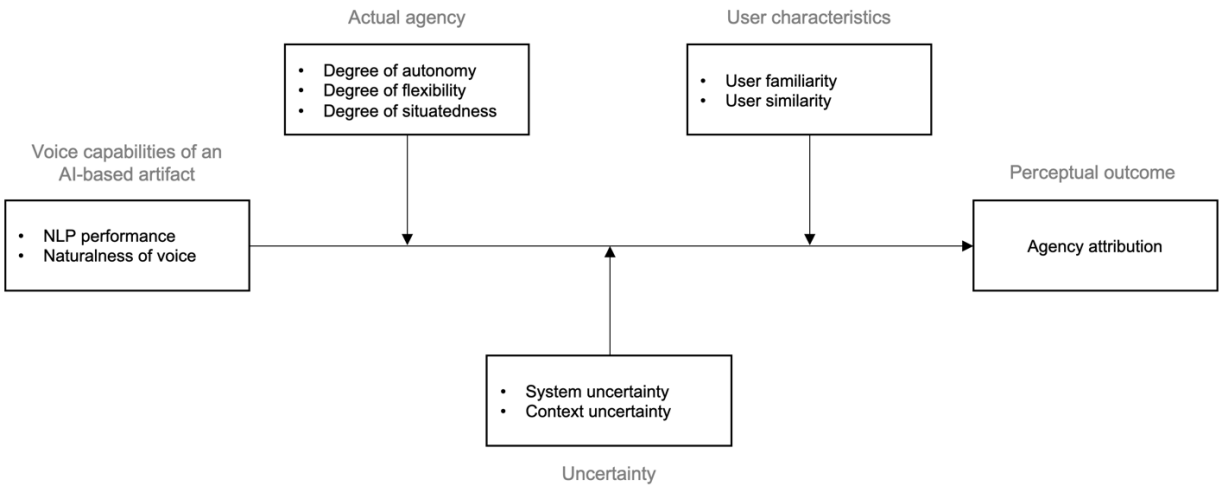

Figure 2. Conceptual Model

3.1 Voice Capabilities

We argue that an artifact's voice capabilities are a key driver in agency attribution and that they exacerbate potentially inappropriate levels of attribution, with such capabilities defined as the machine-based synthesis of voice embedded in AI-based artifacts (Assumption 2) (Figure 2). The main mechanism through which voice capabilities can enhance the perceived agency of an artifact is by interacting with users in a dialogue-based, interactive manner (Balasuriya et al., 2018; Seaborn et al., 2021).

For an AI-based artifact to interact via voice, it must be able to (1) identify, understand and respond to spoken language (hereafter coined as NLP performance) and (2) create a natural voice (hereafter coined as naturalness of voice). The desired emulation capabilities of AI (Assumption 3 of our model) become apparent when examining the design of voice: "The best approach for creating systems that comprehend or produce speech is to model human physiology and cognition. … This approach to the problem of designing speech systems has emerged as the dominant paradigm for design, especially in speech-output systems" (Nass & Gong, 2000, p.38). Both Amazon Alexa and Apple Siri interact with the user by understanding the spoken language of the user and by providing an answer that sounds very natural. Interaction design theories suggest that when an artifact sounds human and follows that pattern of a voice-based interaction between humans, users are inclined to behave and make attributions about the AI-based artifact using the same attribution mechanisms applied to humans (Nass & Gong, 2000).

In line with research on stimulus generalization (Guttman & Kalish, 1956; Shepard, 1987), "organisms extend highly practiced stimulus responses to new stimuli if they resemble the original" (Zhao & Malle, 2022, p. 2). This notion can be applied to comparisons between stimuli deriving from humans versus machines: "Individuals behave toward and make attributions about voice systems using the same rules and heuristics they would normally apply to other humans" (Nass & Gong, 2000, p. 38). Studies have shown that users assign personality traits to synthetic voices, such as Apple Siri or Microsoft Cortana (Hacker, 2021), and even expect voice-based artifacts to exhibit characteristics that are usually attributed to humans (Kim et al., 2021). Amazon Alexa users may consider the voice-based artifact to be their new best friend (Purington et al., 2017) or perceive a robot voice to be "smiling"—an attribution usually made to humans (Torre et al., 2020). Understanding how humans react to fellow humans can be helpful in determining how users will react to AI-based artifacts. Presumably, if users unconsciously interpret the voice of an AI-based artifact as they would a human voice, users would be expected to attribute agency to an artifact much as they would to a human. With increasing ubiquitous computing moving toward hands- and eyes-free computing focusing on voice, this social attribution is exacerbated even further, as no further cues beyond auditory cues may be available to the user (Natale, 2021). For example, Google's Duplex made users believe they are talking to another human, as illustrated by recordings at Google's developer conference I/O (O'Leary, 2019). NLP performance and the naturalness of voice influence users' agency attribution through stimulus generalization (Guttman & Kalish, 1956).

Effectively, both NLP performance and the naturalness of an artifact's voice and their effect on users' cognitive sensemaking (Guttman & Kalish, 1956; Nass & Gong, 2000; Natale, 2021; Shepard, 1987) must be taken into account to understand how AI-based artifacts' voice capabilities impact agency attribution.

3.1.1 Natural Language Processing Performance

Users interacting with an AI-based artifact are free in terms of the input they provide and how they react and use the output provided by the artifact. For example, users interacting with an Amazon Alexa can independently and flexibly decide when and how to pose any questions or demands they wish. Because of users' autonomy, different users might pose the same question using very different words or syntax. In a similar vein, voice-based interactions might be influenced by background noise or a user's accent. Such factors can influence the artifact's ability to understand users' input and thus impede the NLP performance of a voice-based artifact.

NLP is embedded in personal assistants such as Apple Siri, predictive text, language translation, and information retrieval technology. With such technologies being increasingly deployed to assist individuals in their daily tasks and augment their decision-making, NLP enables such assistance and augmentation, allowing it to become even more effortless (Gregory et al., 2021). An artifact's NLP performance can be understood as its ability to understand human language as it is spoken and written. In other words, NLP helps machines communicate with humans in human language. When discussing NLP performance, a variety of performance metrics (i.e., accuracy, precision, recall) enable the measurement of the artifact's capabilities from a technical perspective (Billsus & Pazzani, 1999; Sujatha & Rajagopalan, 2017). Artifacts enabled by high NLP performance capabilities can result in a higher accuracy of understanding a user's input, generating results that achieve the purpose of an interaction in an efficient and skillful manner.

In an ideal scenario, an AI-based artifact would decode natural language as input and encode appropriate natural language as output (Keyser et al., 2019). For instance, naive Bayes classifier models are trained on a certain number of intents with a predefined confidence score. As answers are predicted based on the probability of an identified intent, the confidence score determines whether a user input belongs to a specific intent and hence whether and what kind of answer to provide (Narynov et al., 2021; Ruan et al., 2019). While a higher model confidence might classify answers more accurately, it might also identify intent less reliably, resulting in the artifact being unable to provide an answer or asking users to reformulate their input. A lower model confidence, in turn, might offer a lower barrier to identifying intent but might classify answers less correctly—by providing answers that are not goal oriented or do not fully match the user's intent. In both cases, breakdowns will lead to reduced NLP performance (as measured in accuracy) and, from a user perspective, may dampen the perceived competencies of the AI-based artifact. In a similar vein, large language models might be limited in their performance due to limited model size and data collected in limited contexts (Tamkin et al., 2021). Considering the temporal dimension of voice-based interactions, the same holds true if an AI-based artifact does not properly recognize when a user input is finished, e.g., by not identifying natural pauses and subsequently interrupting the user or posing an unnecessary follow-up question. This can result in interaction breakdowns that make users unsure about how to continue interacting with the artifact and reduce the naturalness of the interaction. Performance failure can thus occur at multiple points, including automatic speech recognition, natural language understanding, and natural language generation (see Figure 1). On this basis, we posit:

Proposition 1a: The greater the natural language processing performance of an AI-based artifact, the higher the agency attributed to this artifact is likely to be.

3.1.2 Naturalness of Voice

As illustrated by the previous examples, the performance of an AI-based artifact in both understanding and reacting via voice has a detrimental impact on the flow of the interaction and the perceptions of the user, which are also influenced by the naturalness of the artifact's voice. The naturalness of a voice can be understood as the replicability of human auditory cues in the design of AI-based artifacts (Assumption 3) (Epley et al., 2007; Seaborn et al., 2021; Zheng & Jarvenpaa, 2021). The ability to resemble the human voice (i.e., giving an artifact a human voice) is enabled through the technical complexity, variety of, and interplay of various auditory cues encodable in AI-based artifacts. According to media equation theory, humans assign personalities not only to other humans but also to machines (Reeves & Nass, 1996), and modifying and combining auditory cues creates certain personality trait associations (Chang et al., 2018). Examining commercially available artifacts including Google Assistant, Microsoft Cortana, and Amazon Alexa, Völkel et al. (2020) demonstrated how artifacts are ascribed certain personality traits that we would normally ascribe to humans and how the perceptions of machine personality can be deliberately shaped. In an e-commerce setting, Nass and Lee (2001) manipulated four auditory cues (volume, fundamental frequency, frequency range, and speech rate) to create extroverted and introverted voice personalities and found that the personality conveyed by the voice had a significant effect on user perceptions: Participants who listened to the same book review rated it differently in terms of liking and credibility depending on the voice. According to the authors, to maximize liking and trust, auditory cues should be set in a way that creates a personality consistent with the user—human—and the presented content.

Commercial developers are exploring the possibilities of auditory cues to deliver a natural experience and improve interaction flow, including Amazon's "one-breath test" to guide the speech rate of a synthetic voice or the design of pauses between a certain number of words (Branham & Roy, 2019; Low, 2020). Other examples of speech design software that manipulate the naturalness of voice include Amazon Polly[1] and Google WaveNet, which enable the direct modification of accent and gender and allow users to train a synthetic voice using samples of their own recorded voice. Commercial artifacts, such as Apple Siri, are increasingly providing alternative gender (and accent) options in addition to the default female voice (Khaled, 2021; Tolmeijer et al., 2021). Interestingly, additional features allow for the direct manipulation of emotions, such as excitement or disappointment (e.g., Amazon Polly) or the creation of a whispering voice (e.g., MaryTTS[1]). Research has demonstrated that the imitation of auditory cues found in human-to-human interactions, such as response latency, can increase the perceived naturalness of artifacts and make up for a lack of logical flow in interactions with AI-based artifacts (Marge et al., 2010).

As these examples illustrate, artifact voice capabilities that enable greater voice naturalness not only induce users to apply heuristics and attributions to AI-based artifacts that are usually applied to other humans but greater artifact voice naturalness also induces higher perceptions of hypothesized general performance. As proposed by the second assumption of our model, designing for social cues further exacerbates the human tendency to apply social rules to machines (Guttman & Kalish, 1956; Puranam & Vanneste, 2022; Shepard, 1987; Zheng & Jarvenpaa, 2021).

Proposition 1b: The greater the naturalness of the voice of an AI-based artifact, the higher the agency attributed to the artifact is likely to be.

3.1.3 Actual Agency

From a sociological perspective, agency is commonly referred to as the ability to "think, plan, and act" (Puranam & Vanneste, 2022, p. 3) or to "act with intent" (Murray et al., 2021, p. 7)—an ability usually ascribed to humans (Assumption 3 within our model). Similarly, Russel and Norvig (2016) define artifacts as possessing agency when they are able to perceive and act on their own. This perspective has been applied by numerous IS scholars, such as Schuetz and Venkatesh (2020), who argue that machines are increasingly incorporating humanlike capabilities, and Baird and Maruping (2021), who theorize about IS artifacts becoming agentic by actively and autonomously executing tasks. We thus suggest that the effect of an artifact's voice capabilities on attributed agency is moderated by the executing capabilities of this artifact. To ensure such a strengthening effect, the expectations evoked by the voice-based interaction must be fulfilled by an artifact's "actual agency," defined as the ability to act toward a goal in an autonomous, situated, and flexible manner (Baird & Maruping, 2021; Gray et al., 2007; Jia et al., 2022; Jennings et al., 1998; Russell & Norvig, 2016).

Increasing artifacts' perceived naturalness through voice can lead to frustration, especially if an artifact's competencies are not ensured. Actual agency acts as a mechanism within voice-based interaction by helping fuel the agency attributed to an AI-based artifact through increased NLP performance and naturalness of the artifact's voice as users get a better idea of an artifact's "degree of intelligence" and its actual capabilities.

To understand the moderating effect of an artifact's actual agency on the relationship between the voice capabilities of an AI-based artifact and agency attribution, consider the interaction contexts discussed earlier. Beyond auditory cues, several factors within an interaction can increase the agency of an artifact. Amazon Alexa, for instance, continuously improves by learning from past interaction breakdowns, considers past interactions and sensory input to proactively start an interaction, and adapts its voice to the interaction context (Low, 2020; Tarantola, 2020). Voice-based artifacts are deployed for a variety of tasks, including general purpose tasks, e.g., automotive user interfaces providing drivers with relevant direction information or making specific suggestions based on comparisons and predictions, i.e., in healthcare or educational settings. However, few AI-based artifacts fully execute tasks themselves, e.g., as we would expect from autonomous driving. Accordingly, we suggest that an artifact's degree of autonomy (i.e., executing tasks autonomously), degree of flexibility (i.e., acting in a responsive, proactive, and social manner), and degree of situatedness (i.e., reacting to sensory input)—which is in line with Jenning's et al.'s (1998) conceptualization of agency—need to be considered as factors moderating the impact of voice capabilities on attributed agency.

3.1.4 Autonomy

Increased NLP performance and voice naturalness—the main mechanisms through which voice capabilities of an artifact positively influence agency attribution—depend on the degree of autonomy an artifact possesses and enacts. Having the ability to control information and make decisions, autonomous artifacts interact without

[1] Amazon Polly is a proprietary TTS system (https://aws.amazon.com/de/polly/) and MaryTTS (https://github.com/marytts/marytts) is an open-source TTS software.

human intervention (Berente et al., 2021; Möhlmann et al., 2021) which becomes apparent in scenarios such as autonomous driving (Frazzoli et al., 2002), investment advisory (Lee & Shin, 2018), and loan processing (Markus, 2017). According to the seminal agency literature, autonomy implies the ability to control one's own actions and leverage prior experiences and the information one has gathered (Emirbayer & Mische, 1998; Jennings et al., 1998; Russell & Norvig, 2016). Turning towards voice-based artifacts, such artifacts can independently react to and generate output in response to a user input yet also draw on past user inputs. For example, according to Google's "tapering" strategy, voice-based artifacts adapt the level of detail provided in an interaction according to the interaction history with a user. If a user has interacted many times with the artifact, the prompts provided become less comprehensive, compared to the more extensive prompts provided to a novice user (Branham & Roy, 2019). The more autonomous the AI-based artifact, the greater the ability to control one's own actions and leverage previous interactions, thereby increasing the interaction capabilities of the voice-based artifact. Berger et al. (2021) illustrate how an artifact's ability to learn and improve increases users' reliance on the artifact's advice. Artifacts inductively improve through data and experience (Berente et al., 2021), as depicted by deep and reinforcement learning (LeCun et al., 2015) and fueled by the surge of big data (Chen et al., 2012). Yet the overall presence of voice-based artifacts executing tasks fully on their own is limited, highlighting again the importance of autonomy as an important moderator of the relationship between an artifact's voice capabilities and its attributed agency.

Proposition 2a. The greater the degree of autonomy of an AI-based artifact, the stronger the relationship between the voice capabilities of an AI-based artifact and the agency attributed to the artifact will be.

3.1.5 Flexibility

Increased NLP performance and voice naturalness also depend on the artifact's degree of flexibility in terms of being able to adapt to an environment in a proactive and social manner (Jennings et al., 1998; Schuetz & Venkatesh, 2020). "Flexibility" encompasses aspects of goal directedness (capability of making plans, being proactive, and working toward a goal), responsiveness (capability of perceiving the environment and responding to changes in it), sociability (capability of interacting with others and supporting them in their activities), and self-control (capability of exercising self-restraint over desires, emotions, and impulses) (Gray et al., 2007; Jennings et al., 1998; Schuetz & Venkatesh, 2020). In their experimental study, Schuetzler et al. (2020), for instance, consider a voice-based artifact's conversational skills in terms of providing varied and tailored responses to assess the artifact's ability to engage with the user in a social manner. The greater the degree of flexibility, the greater the signaling of an artifact's abilities, thereby strengthening the impact of voice capabilities on attributed agency.

Flexibility becomes relevant in that commercially available voice-based artifacts can be tailored toward certain target populations in terms of the modifiability of the speech rate and the time-out periods of the voice output. For example, an artifact can increase its rate of speech when interacting with visually impaired users (who can usually process voice considerably faster than users without impaired vision), and longer voice time-out periods for elderly or disabled people who may take longer to formulate a command (Branham & Roy, 2019). Such flexibility not only offers greater personalization and interactions that can fulfill the needs and preferences of individual users, but they can also provide minorities with new technological interaction possibilities and reach neglected user groups (Metatla et al., 2019; Schlögl et al., 2013; Sutton et al., 2019). Another example is that of deaf or hearing-impaired individuals who require customization options. Default voice-based artifacts often use a high pitch, which may be incompatible with hearing aids (Blair & Abdullah, 2020). The more goal-directed, responsive, social, and controlled the activities of a voice-based artifact (e.g., responsiveness to users' individual and contextual factors), the more natural and performative voice-based interaction becomes, making the artifact appear to be more competent.

Proposition 2b. The greater the degree of flexibility of an AI-based artifact, the stronger the relationship between the voice capabilities of the AI-based artifact and the agency attributed to the artifact will be.

3.1.6 Situatedness

The moderating effect of the agency of an AI-based artifact on the relationship between an artifact's voice capabilities and attributed agency also depends on the artifact's degree of situatedness, defined as the artifact's ability to receive, react to, and influence sensors in an environment (Baird & Maruping, 2021; Jennings et al., 1998). Schuetz and Venkatesh (2020) refer to this ability as contextuality (p. 463): "[Artifacts] may draw on multiple sources of information, including both structured and unstructured digital information, as well as sensory inputs (visual, gestural, auditory, or sensor-provided)." This implies that the user does not have to act as a middleman but interacts directly with the external, contextual environment. In this sense, some decisions and input criteria might be opaque to the user. For instance, a voice-based artifact operating in a medical surgery room monitors the real-time environment and performs user-independent actions to modify the environment by being connected with other sensory devices, i.e., by adjusting an operating table or

increasing gas pressure (Perrakis et al., 2012). This strategy can be extended to other aspects, i.e., an artifact's voice becoming louder when the background noise of a user's environment increases.

AI-based artifacts exhibit smart material properties that interfaces in ecosystems can easily utilize—for instance, when context factors trigger ecosystem orchestrators such as IFTTT[2] services (Knote et al., 2021). Amazon Alexa, for instance, provides this situatedness through built-in routines such as smart home activities, including adjusting the lightning or room temperature, reminding users to drink water when coughing is detected, and playing soothing sounds to calm down users' dogs when barking is detected. IFTTT scenarios are also relevant in the organizational context, i.e., when project management applications and documents are connected to a voice-based artifact (Assumption 1 in our model). A financial analyst could, for instance, ask Alexa to update or modify a spreadsheet. In a similar vein, a manager could ask Alexa to add a "to do," which then leads the voice-based artifact to update the manager's calendar (Finnegan, 2017). While everyday voice-based artifacts are increasingly connected with other devices, many actions are still controlled via the voice input of the human user, pointing towards the importance of artifact situatedness as an important moderator of the relationship between the voice capabilities of an AI-based artifact and agency attribution.

Proposition 2c: The greater the degree of the situatedness of an AI-based artifact, the stronger the relationship between the voice capabilities of the AI-based artifact and the agency attributed to the artifact will be.

3.2 Uncertainty

Users of voice-based artifacts often interact with them in contexts of uncertainty, although users simply relying on these artifacts can involve high risks and high personal costs in cases of system error (Assumption 1 in our model). In such interaction contexts, the auditory cues of the artifact are often the only cues available to the user to make sense of a (decision-making) situation (Assumption 2 in our model), making the artifact low-defined from the user perspective (Natale, 2021). While research suggests that transparency and explainability are important design measures that can help users make sense of the technical nature and underlying working mechanisms of AI-based artifacts (Berente et al., 2021), we rarely find these in commercially available voice-based artifacts.

Uncertainty, defined as the degree of ambiguity that users experience in an interaction with an AI-based artifact in a given context (Melara & Mounts, 1994; Pavlou et al., 2007), acts as a key determinant driving human cognition. Underlying system information, including the logic of the predictive model, the data it was trained on, the performance of the model (Knight, 2017), and details on the providers and institutional factors of an artifact (Granados et al., 2010; Vimalkumar et al., 2021) could help users better estimate the competencies of an artifact in a specific context. However, this information is typically not disclosed to users. Accordingly, when this information is unavailable, users must rely on the cues available to them and use their own heuristics to make sense of an interaction context and the party with which they are interacting.

Principal-agent theory helps explain how uncertainty induces users to attribute human characteristics to make sense of an artifact's actions (Pavlou et al., 2007) and "increases confidence in predictions of this agent in the future" (Epley et al., 2007, p. 866). These kinds of heuristics are important for agency attribution because they provide critical drivers regardless of the actual agency of the system on which the use of the artifact depends. We differentiate between two sources of uncertainty that exacerbate agency attribution: (1) how the AI-based artifact itself is designed and disclosed, i.e., the outcome and effects of an artifact's actions or the underlying model of an artifact are uncertain, and (2) how familiar the user is with the context in which they are interacting with the AI-based artifact and the environmental cues available in this context. We argue that uncertainty, both in terms of system and context, must be considered in our model, as it may play an important role in exacerbating the relationship between an artifact's voice capabilities and agency attribution by clouding users' understanding of the artifact and inducing users to attribute levels of agency that are greater than they are necessarily experiencing.

3.2.1 System Uncertainty

Uncertainty can arise due to missing information when trying to make sense of an AI-based artifact, i.e., the imperfect information and opaque nature of AI-based artifacts may make it difficult to accurately predict an artifact's true attributes and competencies (Berente et al., 2021; Pfeffer & Salancik, 1978). As proposed by Berente et al. (2021, p. 22), the degree of transparency represents "the amount [artifact] owners wish to disclose or occlude" about the artifact at hand. First, given the stochastic nature of ML-based artifacts, the

[2] IFTTT is an abbreviation for "if this, then that" and is a service to orchestrate various IT artifacts based upon several (contextual) factors. Examples include "If you add a new task to your Amazon Alexa to-dos, then it will be added to your iOS reminder app" or "If the International Space Station passes over your house, then you'll get a smartphone notification about it" (Martin & Finnegan, 2020)

underlying logic of a model and its respective output cannot be fully predicted and will always lead to a limited amount of visibility into the system (Knight, 2017). Second, the intransparent design of AI-based artifacts, particularly regarding voice, may facilitate or cause deception (Sarkadi et al., 2021) when artifacts are mistaken for humans (Gehl & Bakardijeva, 2016) and thereby mobilize users' tendency to attribute agency (Natale, 2021). Google Duplex, for instance, is an extension of Google Assistant, which has demonstrated the capacity to take over service tasks previously executed by humans, e.g., taking care of dinner bookings or hairdressing reservations via phone. In demonstrating this capacity, Duplex did not disclose itself as not being human nor identify the technical processes it was using, which placed the presented technology in the limelight of media discussions and invoked social concerns (O'Leary, 2019). Schuetz and Venkatesh (2020) highlight this issue by challenging the prevailing assumption that humans are always aware that they are interacting with an AI-based artifact.

According to Natale (2021), users both deliberately and unconsciously engage with AI-based artifacts, reaping their pragmatic benefits without having to deal with the underlying technical complexities associated with such artifacts—a phenomenon coined as "banal deception." A common example of such engagement is the AI-based artifact Replika, which mimics users' texting style and offers 24/7 companionship to its users. As voice-based artifacts are not necessarily embodied, higher levels of deception can be induced. Anthropomorphism acts as a coping mechanism to reduce system uncertainty (Epley et al., 2007; Pavlou et al., 2007) in that "voice assistants encourage users, by relying on them to apply their own stereotyping, to contribute actively to the construction of sense around the disembodied voice" (Natale, 2021, p. 114). Doing so, however, induces users to uncritically and incompletely engage with the artifact at hand, ultimately leading users to ascribe capabilities to the system out of convenience.

Proposition 3a: The greater the uncertainty a user associates with the artifact they are interacting with, the stronger the relationship between the voice capabilities of an AI-based artifact and the agency attributed to the artifact will be.

3.2.2 Context Uncertainty

Uncertainty can also arise due to missing information in order to make sense of an interaction context, i.e., driving in an unknown city or making a forecasting decision with limited information on the forecasting task and topic. In that sense, contextual uncertainty encompasses the degree to which the (future) state of the environment cannot be accurately anticipated or predicted due to imperfect information and ambiguity regarding how to arrive at an ideal future state (Pfeffer & Salancik, 1978).

Similar to system uncertainty, context uncertainty can induce users to attempt to reduce or cope with the uncertainty in particular ways, i.e., by relying on the system. For instance, news outlets in a number of countries have regularly reported on cases in which drivers blindly followed the instructions of their navigation systems, which ultimately led to dangerous and even fatal outcomes for the users of these systems (Frankfurter Allgemeine Zeitung, 2022; Lin et al., 2017; Nederlandse Omrop Stichting, 2020). This illustrates the potential pitfalls of users blindly relying on such systems when they find themselves in uncertain situations, such as driving in a foreign country. In a similar vein, Logg et al. (2019) found that in a forecasting task where users had little to no expertise in the forecasting decision at hand, they were more likely to rely on the advice given by an AI-based artifact. Both examples demonstrate that when contextual uncertainty arises, it becomes difficult for a user to assess potential risks and an appropriate decision (Baier, 1986; Jones, 1996). At the same time, effectance motivation assumes that humans aim to interact effectively and thus reduce uncertainty (Epley et al., 2007). Attributing agency to an artifact can therefore function as a coping mechanism to reduce contextual uncertainty and increase comprehensibility in the user's environment.

Proposition 3b: The greater the uncertainty a user associates with the context of an interaction, the stronger the relationship between the voice capabilities of an AI-based artifact and the agency attributed to the artifact will be.

3.3 User Characteristics

The attributed agency of an AI-based artifact is only as strong as the users' capability to understand and assess the artifact's competencies. The more familiar a user is with voice-based artifacts and the more an artifact's voice resonates with the user, the stronger the agency attributed to the artifact is likely to be. With everyday voice-based interactions increasingly being introduced to organizational decision-making (Assumption 1 in our model), IT consumerization stresses the importance of considering users' previous interactions and identification with an artifact. Berente et al. (2021) coined this notion of AI-based artifacts as "inscrutability," with such artifacts being "intelligible only to a select audience while remaining opaque to others, or, in some cases, not intelligible to humans at all" (p. 1437).

User characteristics involve individual user differences such as a user's propensities and dispositions, demographic aspects, cultural background, and experience (Zhang & Li, 2005; Zhang et al., 2002). Understanding certain user characteristics is crucial, as "individual characteristics often influence how [users] perceive the outcome of an event and attribute responsibility, and their interaction with technology"

(Jia et al., 2022, p. 390). For instance, research has shown that voice-based artifacts are perceived differently depending on the user's gender (Song et al., 2020). We introduce Brunswik's (1956) notion of the "lens" to the information systems community to explain the human judgment process and how it is influenced by individual factors, as compared to a normative judgment process. The sensemaking of a voice-based artifact and its output is thus based on the perspectives and knowledge of the user. The lens perspective provides insights into how an environmental stimulus and its variables (in this case, an AI-based artifact's voice and auditory cues) influence an individual's (in this case, an artifact's user's) perceptions (Scholz, 2017). The perceptions of a user are not necessarily inscribed in individual, distinguishable auditory cues but must be inferred by combining and interpreting individual cues and by taking environmental factors such as user characteristics into account. In this sense, users rely on their own lens and make sense of interaction contexts based on familiar cues and experiences.

Thus, while certain individual and combined auditory cues evoke certain associations, caution is required when attempting to generalize because each user acquires and interprets these cues through their own contextual lens. Therefore, user characteristics act as another key mechanism of voice-based interaction with AI-based artifacts. We suggest that both the familiarity and the similarity of a user with an AI-based artifact need to be considered as factors moderating the impact of voice capabilities on agency attribution in that they influence how users interpret and interact with an artifact.

3.3.1 User Familiarity

To fully understand why and to what extent users attribute agency to voice-based artifacts, it is necessary to assess the extent to which the user's cognitive process incorporates considerations of familiarity. User familiarity is defined as the extent of a user's previous experiences with, interactions with, and learning regarding similar artifacts (Komiak & Benbasat, 2006; Rotter, 1971). Based on Assumption 1 of our model, we view the interaction with an AI-based artifact not as completely new but rather one that is usually informed by comparable experiences, i.e., interaction with the same or other types of voice-based artifacts. According to Rotter (1971), previous experiences with situations determine expectations for a particular situation that is perceived to be similar. Familiarity is thus acquired through a user's prior and direct experiential exchanges with an artifact. Hence, the understanding and sensemaking of an AI-based artifact are a function of the user's literacy regarding such artifacts (Kovalerchuk et al., 2021).

The more a user learns to express their needs in an interaction and learns what types of questions to ask and what kinds of answers or explanations an artifact can give, the more familiarity with the artifact increases (Komiak & Benbasat, 2006). The more familiar users become with an artifact, the more their expectations regarding the actual capabilities of an artifact will become calibrated and the more they will be able to assess such capabilities. Accordingly, user familiarity will likely strengthen the impact of voice capabilities on agency attribution. For instance, consider a scenario in which a user attempts to optimize their smart home by configuring and using as many smart devices as possible. A user who is already very familiar with Amazon Alexa, knows Alexa's "wake" words, and understands how to set up a new, networked smart device, will experience Alexa as much more agentic than a novice Alexa user. Since users are encouraged to use Alexa multiple times a day, even though they are capable of, for example, turning off their lights themselves, the AI-based artifact continuously supplies the underlying voice capabilities with new user input data, helping it improve its voice-based interaction with the user. Many companies such as Amazon are taking advantage of this effect by providing users with detailed information about an artifact's capabilities prior to the first interaction (often, the artifact itself communicates this information) to prevent situations in which the limits of an artifact's competencies are challenged. As a result, system designers can help increase the agency attributed to it by establishing and increasing users' literacy regarding the artifact.

Proposition 4a: The greater the degree of familiarity of a user with an AI-based artifact, the stronger the relationship between the voice capabilities of the AI-based artifact and the agency attributed to the artifact will be.

3.3.2 User Similarity

The rationale behind agency attribution to a voice-based artifact is also a function of user similarity, defined as the perceived similarity between an artifact's voice and the user's voice in terms of auditory cues (Byrne, 1971). Analogous to user familiarity, user similarity also shapes users' attribution of agency to an artifact and, by extension, the level of trust in the system. As humans generally seek to affiliate with or be in the presence of others (Cheek & Buss, 1981), similarity with an artifact's voice allows the user to identify with the artifact (Suh et al., 2011). In fact, "similarity to self …, provides evidence that one is functioning in a logical and meaningful manner" (Byrne et al., 1967, p. 83), which leads to an attribution of naturalness, preference, and trust (e.g., Cowan et al., 2016; Dahlbäck et al., 2007).

Agency attribution is crucial in trust-relevant contexts such as automated driving and thus requires a voice with which the user can identify. Truschin et al. (2014) illustrated how matching the gender of the artifact's voice to the user's gender reduces cognitive demand in driving scenarios, ultimately making the interaction with the voice-based artifact more effective. In another real-world driving study, Braun et al. (2019) found that

matching the word choice and intonation of the automotive user interface to that of the driver's caused the interface to be perceived as more trustworthy and likable than a default voice character. Building on similarity attraction theory, further empirical studies have demonstrated that desirable interaction outcomes can be induced through similarity between the artifact and the user, which would thus require the artifact's voice to be adapted to the human voice of the user (Byrne, 1971). As such, Eyssel et al. (2012) reported that same-gender voice-based robots were perceived more positively by the user, were associated with greater psychological closeness, and were more strongly anthropomorphized. In a similar vein, matching an AI-based artifact's accent to that of the user increases user acceptance (Cowan et al., 2016). Commercial providers have realized the implications of mirroring the user's voice with more personalized options in terms of language and dialect—e.g., Apple offering gender/dialect choices for Siri. High similarity between the user and artifact can help users ascribe positive competencies to the artifact, including the agency attributed to the artifact.

Proposition 4b: The greater the degree of similarity between a user and an AI-based artifact, the stronger the relationship between the voice capabilities of the AI-based artifact and the agency attributed to the artifact will be.

4 Discussion

Our research contributes to the literature on user interaction with AI-based artifacts (Benbya et al., 2021; Diederich et al., 2022, Murtarelli et al., 2021; Puranam & Vanneste, 2022; Schuetz & Venkatesh, 2020) by explaining the role of artifacts' voice capabilities for agency attribution. In the context of AI-based artifacts, agency attribution is particularly interesting for voice-based interactions due to the modality's unique properties and interaction implications. We posit a positive direct relationship between the voice capabilities of an AI-based artifact and the agency attributed to the artifact by its users—a relationship that is moderated by the artifact's actual agency, the uncertainty revolving around users' interactions with the artifact, and users' characteristics. We shed light on new and increasingly prevalent forms of interaction with AI-based artifacts, wherein a user's perception of the agency of an artifact is a function of the voice capabilities of the artifact. Integrating our model of voice-based interaction with AI-based artifacts with the extant agency literature, we argue that the agency attributed to AI-based artifacts is made stronger through the use of voice as an interaction modality and is realized through the naturalness of the artifact's voice as well as its natural language performance. This highlights the need to examine the key implications and boundaries of the conceptual model and introduces future research avenues.

Our explanation of agency attribution in the era of AI-assisted decision-making offers a novel set of insights. First, prior research describes an increasingly crucial and distinctive category of interaction with AI-based artifacts focused on voice-based interaction. As stated by Gregory et al. (2021), voice-based artifacts offer a "low effort interaction with AI for users and thus fuel sustainable adoption and use as they offer a more natural and intuitive way for humans to interact with the machines and use products and services" (p. 543). In such contexts, the user experience is heavily shaped by the naturalness of the artifact's voice and the cues available in the artifact. We surmise that voice-based interaction largely influences a user's overall perception of an AI-based artifact and subsequent behavioral consequences, i.e., decision-making (Baird & Maruping, 2021; Berente et al., 2021; Brynjolfsson, 2022; Gregory et al., 2021; Metcalf et al., 2019; Raisch & Krakowski, 2020).

Previously introduced opportunities around the richness of the voice modality also present new or existing accessibility, privacy, and personal data challenges. Due to the nonverbal cues transmitted in a voice-based message, more privacy-sensitive information about the users themselves is conveyed (e.g., gender, age, emotional state). Further, voice-based interactions may induce users to more easily disclose sensitive information due to the natural and effortless type of interaction. As NLP performance hinges on the richness and volume of user input, the collection of user input puts the user at greater risk in that voice can serve as a biometric identifier. Security and privacy issues should thus be more thoroughly addressed in future research. As noted by Lowry et al. (2017), questions regarding user compliance, data collection, and aggregation have yet to be discussed in the context of voice-based interaction with AI-based artifacts. For instance, Dickhaut et al. (in press) discussed the storage of voice-based user data on the device versus in the cloud regarding its privacy and system performance implications. This becomes particularly important when considering increasing IT consumerization (Assumption 1 of our conceptual model), where the lines of personal and work-related IT use are increasingly blurred.

Second, our conceptual model demonstrates how inappropriate levels of agency attribution might be induced through the voice capabilities of an AI-based artifact alone. For instance, having a high NLP performance does not per se lead to agentic artifacts. Large language models perform well in generating natural language output that is sophisticated and correct in terms of syntax and grammar. However, this output is often unfaithful or nonsensical, e.g., when it refers to a nonexistent information source. Despite a lack of contextualization and quality, i.e., abilities of reasoning and understanding, high NLP performance, i.e., linguistic capacities, can make a system be perceived as

highly sophisticated, competent—and agentic. Researchers have thus started to warn about the linguistic capabilities of such models as "hallucinations" and deceitful, as users might thereby perceive an AI-based artifact's overall agency to be wrongfully high (Ji et al., 2023). While such models are largely deployed in text-based interaction, recent developments such as GPT-4 including auditory input make the large-scale deployment of large language models in voice-based interaction likely.

In addition, the CASA paradigm and social response theory are the dominant theoretical lenses guiding the design of AI-based artifacts and also apply to voice-based artifacts (Schmitt et al., 2021). Voice design theory tends to apply heuristics known from interpersonal communication to interaction with AI-based artifacts, with voices imitating the human voice. Previous studies have explored personality models for agents and interfaces (Völkel et al., 2020) or built their theorizing on the notion that perceptions of the human voice also apply to perceptions of artifacts' voices (Marge et al., 2010). At the same time, known models of human-human communication are often inappropriately applied to HCI and design strategies for modality-agnostic HCI are not necessarily suitable for voice-based interaction (Guzman, 2019). For instance, non-humanlike spoken dialogue design has been shown to reduce speech collisions during interactions with a voice-based artifact (Funakoshi et al., 2010). From an ethical perspective, relying on the notion that an AI-based artifact's voice should resemble the human voice—for example, through increased naturalness—can be a hazardous strategy. Anthropomorphizing a voice-based artifact in its design may increase its attributed agency, potentially beyond its technical capabilities, ultimately leading to overreliance or mistrust (Puranam & Vanneste, 2022). Such design strategies raise the question of how to calibrate the "naturalness" of a voice-based artifact to ensure levels of trust, reliance, and user behavior that match the artifact's actual agency.

Future research could complement the body of HCI research focusing on system capabilities and users' understanding of and conscious engagement with AI-based artifacts. This research stream is concerned with, for instance, whether and how objective information on an artifact's performance and capabilities or evaluative decision explanations might induce appropriate reliance in AI-assisted decision-making (Bansal et al., 2019; Lu & Yin, 2021; Miller, 2023; Zhang et al., 2020). Voice-related design choices, i.e., the use of voice as an interaction modality and its design could hence be guided by when what levels of agency attribution are warranted—and when they are not. In line with the idea of reducing system uncertainty, design studies could explore when and how to disclose AI-based artifacts interacting via voice as nonhuman, or in which cases it is desirable to reduce the naturalness of an artifact's voice. Moving beyond our conceptual model, we suggest empirically examining the impact of AI-based artifacts and voice-based interaction on the downstream consequences of agency attribution, including user trust and behavioral outcomes such as decision-making quality (Puranam & Vanneste, 2022). The resulting attribution may manifest in trust (Behrens et al., 2018; Chiou et al., 2020) or other implications such as reliance on AI advice or increased use (e.g., Cho et al., 2019). Preliminary research has investigated how the modification of auditory cues can promote the trust in and acceptance of voice-based artifacts and help users behave in a socially acceptable and accountable manner (Vance et al., 2015). Mixed results have been reported regarding dependent variables such as perceptions of trust and agency (Hess et al., 2009; Elkins & Derrick, 2013) and it should be questioned whether and how agency attribution acts as an important mediator helping to explain these results.

Third, revisiting the assumptions made in the theory building of agency attribution through voice-based interaction offers important boundary conditions to understanding our conceptual model and raises questions to be addressed in future research. Our focus on voice as a predominant modality and the cognitive user process also implies that agency attribution is instance specific. Research on other perceptual user outcomes such as attributed trust and performance has proposed that it is difficult to generalize user attributions across (dissimilar) tasks and contexts, as expectations and affordances differ from task to task (Mayer et al., 1995; Sitkin & Roth, 1993). Therefore, levels of agency attribution are comparable only across similar tasks and task complexities. These assumptions suggest that interactions where users interact with AI-based artifacts via alternative or multiple modalities might not offer the same propositions for agency attribution. They might also encompass interactions with very different levels of context complexity and task requirements; thus, our understanding of uncertainty and an artifact's actual agency might be inappropriate.

However, AI-based artifacts are composed of multiple modalities—for instance, there are stand-alone voice artifacts, including automotive user interfaces, as well as embodied conversational systems, such as virtual assistants or physical robots (Craig & Schroeder, 2017; Guzman, 2019; Masina et al., 2020; Schöbel et al., in press). Prevalent themes emerging from the literature include comparisons between modalities and investigations of how the presence (or absence) of voice affects the perceptions of, responses to, and interactions with multimodal artifacts. While voice-only interactions enable more focused interaction, i.e., by facilitating learning outcomes such as memorization (Berry et al., 2015), in certain contexts, i.e., retailing and education, voice as an additional interaction modality can make the

interaction unnecessarily complex and can thus potentially modify system uncertainty or user familiarity (Pagani et al., 2019). Relatedly, increased task complexity introduced through the type of task (e.g., information retrieval versus judgment), processing requirements or (number of) actions involved (e.g., IFTTT) might afford the need for multiple modalities and change users' cognitive processes (Campbell, 1988; Wood, 1986). An ethnographic study of elderly people using personal assistants at home found an affinity toward text-based communication (Schlögl et al., 2013). Complementary modalities are crucial in this case since visual notifications can provide meaningful information beyond auditory cues. Related to an artifact's degree of situatedness, users should be able to connect it to additional devices, such as smartphones operating via multiple modalities. Our focus on human interaction with AI-based artifacts using voice as the predominant interaction modality implies that our theory cannot, at present, speak to these interaction cases or to agency from technical or ontological perspectives.

Fourth, our theory assumes that humans' response to new stimuli (i.e., an AI-based artifact) is based on their response to known stimuli (i.e., fellow humans) (Guttman & Kalish, 1956; Shepard, 1987; Zhao & Malle, 2022) and that the agency attributable to an AI-based artifact is comparable to the agency attributable to a human (Brynjolfsson & Mitchell, 2017; Puranam & Vanneste, 2022; Samuel, 1959). Viewing our conceptual model through such a lens questions whether humans and artifacts should be analyzed with the same conceptual apparatus, i.e., agency. We acknowledge that it is our responsibility to distinguish between humans and technology. We do not attempt to deny such differences. Rather, our aim is to understand the sociotechnical nature of information systems and explain users' cognitive processes whereby agency is attributed to technology or, alternatively, to examine the unintended implications that arise through stimulus generalization with AI-based artifacts interacting via voice. We argue that such attribution induced through voice can be hazardous as it potentially fails to account for AI-based artifacts' actual capabilities and deployment opportunities. In that sense, the goal of the study is not to suggest how a user can be induced to ascribe more agency to an artifact but to demonstrate that the cognitive process of agency attribution should be consciously considered in the design and deployment of AI-based artifacts in order to ensure that human augmentation through AI-based artifacts is warranted.

5 Implications for Managers and System Designers

Managers and system designers are well aware that AI-based artifacts, including those interacting via voice, will become increasingly prevalent—regardless of whether the consideration of such artifacts is enforced by the organization or informally introduced through IT consumerization and the employees themselves. Viewing the use of voice-based artifacts to augment human decision-making as a point of departure, decision makers should pay careful attention to three key mechanisms of voice-based interaction with AI-based artifacts: (1) an artifact's actual agency, (2) the context and system uncertainty accompanying voice-based artifacts, and (3) user characteristics. Regarding an artifact's actual agency, this means ensuring that the competencies of an artifact offer appropriate levels of autonomy, flexibility, and situatedness for the task and context in which the artifact is deployed. In terms of uncertainty, this means considering potential ambiguities that users may encounter in their decision-making processes and hence providing individuals with the necessary amount of information needed for a decision-making task. In addition, it is crucial to consider disclosing the AI-based artifact as such and providing complementary information on the artifact (i.e., underlying technical model, model training data, tasks for which the artifact is deployed, artifact performance metrics), which might help the user better estimate the competencies of an artifact and its suitability for decision-making. With respect to user characteristics, this means accounting for users' extant familiarity with similar types of artifacts and potentially offering supplementary training. By considering user characteristics such as gender, age, and visual or hearing impairment the personalization opportunities of voice-based artifacts and individual auditory cues can be leveraged to better cater to the needs of individual users. Bias and stereotypes associated with certain genders and sociocultural backgrounds that could potentially be introduced through voice should be considered in the design of AI-based artifacts as well. When successfully considering and implementing these mechanisms, users will likely attribute calibrated levels of agency to voice-based artifacts, which can then facilitate the sustainable and effective use of AI-based artifacts for human decision-making.

Acknowledgments

We sincerely thank the senior editor Robert Wayne Gregory and the three anonymous reviewers for their constructive guidance and thoughtful feedback on developing the paper. We are also grateful for the opportunity to present an earlier version of this paper at Harvard University's Intelligent Interactive Systems Lab. Thank you Melanie Schwede, Pietro Alessandro Aluffi, and Benjamin Ruben Schmierer for valuable comments on earlier versions of this paper. We thank the Swiss National Science Foundation for funding parts of this research (192718). The third author acknowledges funding from the Basic Research Fund (GFF) of the University of St.Gallen.

References

Abdolrahmani, A., Kuber, R., & Branham, S. M. (2018). Siri talks at you: An empirical investigation of voice-activated personal assistant (VAPA) usage by individuals who are blind. *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 249-258).

Aeschlimann, S., Bleiker, M., Wechner, M., & Gampe, A. (2020). Communicative and social consequences of interactions with voice assistants. *Computers in Human Behavior*, *112*, Article 106466.

Amazon Web Services (2022). *Alexa for business.* https://aws.amazon.com/de/alexaforbusiness/

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*(4), 496-515.

Baier, V. E., March, J. G., & Saetren, H. (1986). Implementation and ambiguity. *Scandinavian Journal of Management Studies*, *2*(3-4), 197-212.

Baird, A., & Maruping, L. M. (2021). The next generation of research on is use: A theoretical framework of delegation to and from agentic is artifacts. *MIS Quarterly*, *45*(1), 315-341.

Balasuriya, S. S., Sitbon, L., Bayor, A. A., Hoogstrate, M., & Brereton, M. (2018). Use of voice activated interfaces by people with intellectual disability. *ACM International Conference Proceeding Series* (pp. 102-112).

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019). Beyond accuracy: The role of mental models in human-AI team performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*.

Behrens, S. I., Egsvang, A. K. K., Hansen, M., & Møllegrd-Schroll, A. M. (2018). Gendered robot voices and their influence on trust. *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (pp. 63-64).

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129-135.

Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: artificial intelligence in organizations: implications for information systems research. *Journal of the Association for Information Systems*, *22*(2), 281-303

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433-1450.

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch me improve—algorithm aversion and demonstrating the ability to learn. *Business & Information Systems Engineering*, *63*(1), 55-68.

Berry, D. C., Butler, L. T., & De Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human Computer Studies*, *63*(3), 304-327.

Billsus, D., & Pazzani, M. J. (1999). A personal news agent that talks, learns and explains. *Proceedings of the Third Annual Conference on Autonomous Agents* (pp. 268-275).

Bird, S., Loper, E. & Klein, E. (2009). *Natural language processing with Python.* O'Reilly Media Inc.

Blair, J., & Abdullah, S. (2020). It didn't sound good with my cochlear implants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(4). Article 118.

Branham, S. M., & Roy, A. R. M. (2019). Reading between the guidelines: How commercial voice assistant guidelines hinder accessibility for blind users. *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 446-458).

Braun, M., Mainz, A., Chadowitz, R., Pfleging, B., & Alt, F. (2019). At your service: Designing voice assistant personalities to improve automotive user interfaces a real world driving study. *Proceedings of the Conference on Human Factors in Computing Systems*.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). University of California Press.

Brynjolfsson, E. (2022). The Turing trap: The promise & peril of human-like artificial intelligence. *Daedalus*, *151*(2), 272-287.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530-1534.

Byrne, D., Griffitt, W., & Stefaniak, D. (1967). Attraction and similarity of personality characteristics. *Journal of Personality and social Psychology*, *5*(1), 82-90.

Byrne, D., Gouaux, C., Griffitt, W., Lamberth, J., Murakawa, N. B. P. M., Prasad, M., ... & Ramirez III, M. (1971). The ubiquitous relationship: Attitude similarity and attraction:

A cross-cultural study. *Human Relations, 24*(3), 201-207.

Cambre, J., & Kulkarni, C. (2019). One voice fits all? Social implications and research challenges of designing voices for smart devices. *Proceedings of the ACM on Human-Computer Interaction*.

Campbell, D. J. (1988). Task complexity: A review and analysis. *The Academy of Management Review, 13*(1), 40-52.

Carlotto, T., & Jaques, P. A. (2016). The effects of animated pedagogical agents in an English-as-a-foreign-language learning environment. *International Journal of Human Computer Studies*, *95*, 15-26.

Carpenter, J., Davis, J. M., Erwin-Stewart, N., Lee, T. R., Bransford, J. D., & Vye, N. (2009). Gender representation and humanoid robots designed for domestic use. *International Journal of Social Robotics, 1*, 261-265.

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

Cheek, J. M., & Buss, A. H. (1981). Shyness and sociability. *Journal of Personality and Social Psychology*, *41*(2), 330-339.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4) 1165-1188.

Chiou, E. K., Schroeder, N. L., & Craig, S. D. (2020). How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers and Education*, *146*, Article 103756.

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

Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). *Voice user interface design*. Addison-Wesley Professional.

Cowan, B. R., Gannon, D., Walsh, J., Kinneen, J., O'keefe, E., & Xie, L. (2016). Towards understanding how speech output affects navigation system credibility. *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 2805-2812).

Craig, S. D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers and Education*, *114*, 193-205.

Crumpton, J., & Bethel, C. L. (2016). A Survey of Using Vocal Prosody to Convey Emotion in Robot Speech. *International Journal of Social Robotics*, *8*(2), 271-285.

Daft, R., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554-571.

Daft, R., Lengel, R., & Trevino, L. K. (1987). Message equivocality, media selection, and manager performance: Implications for information systems. *MIS Quarterly*, *11*(3) 355-366.

Dahlbäck, N., Wang, Q. Y., Nass, C., & Alwin, J. (2007). Similarity is More Important than Expertise: Accent Effects in Speech Interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*

Dennis, A. R., Fuller, R. M., & Valachich, J. S. (2008). Media, Tasks, and Communication Processes: A Theory of Media Synchronicity. *MIS Quarterly*, *32*(3), 575-600.

Dess, G. G., & Beard, D. W. (1984). Dimensions of organizational task environments. *Administrative Science Quarterly*, *29*(1), 52-73.

Dickhaut, E., Janson, A., Söllner, M., & Leimeister, J. M. (in press). Lawfulness by design-development and evaluation of lawful design patterns to consider legal requirements. *European Journal of Information Systems*, https://doi.org/10.1080/0960085X.2023.2174050

Elkins, A. C., & Derrick, D. C. (2013). The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents. *Group Decision and Negotiation*, *22*(5), 897-913.

Elshan, E., Ebel, P., Söllner, M., & Leimeister, J. M. (2023). Leveraging Low Code Development of Smart Personal Assistants: An Integrated Design Approach with the SPADE Method. *Journal of Management Information Systems, 40*(1), 96-129.

Emirbayer, M., & Mische, A. (1998). What is agency?. *American Journal of Sociology, 103*(4), 962-1023.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864-886.

Eyssel, F., Kuchenbrandt, D., Bobinger, S., De Ruiter, L., & Hegel, F. (2012). "If you sound like me, you must be more human": On the interplay of robot and user features on human-robot acceptance and anthropomorphism. *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 125-126).

Frankfurter Allgemeine Zeitung (2022, August 19). *Mann fährt mit Auto auf Freilichtbühne* [Man drives car onto open air stage]. https://www.faz.net/aktuell/gesellschaft/navi-leitet-fehl-mann-faehrt-mit-auto-auf-freilichtbuehne-18254884.html

Frazzoli, E., Dahleh, M. A., & Feron, E. (2002). Real-time motion planning for agile autonomous vehicles. *Journal of Guidance, Control, And Dynamics*, *25*(1), 116-129.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *MIS Quarterly*, *45*(3), 1527-1557.

Funakoshi, K., Nakano, M., Kobayashi, K., Komatsu, T., & Yamada, S. (2010). Non-humanlike spoken dialogue: A design perspective. *Proceedings of the SIGDIAL 2010 Confeence: 11th Annual Meeting of the Special Interest Group On Discourse and Dialogue* (pp. 176-184).

Gálvez, R. H., Gravano, A., Beňuš, Š., Levitan, R., Trnka, M., & Hirschberg, J. (2020). An empirical study of the effect of acoustic-prosodic entrainment on the perceived trustworthiness of conversational avatars. *Speech Communication*, *124*, 46-67.

Gehl, R. W., & Bakardjieva, M. (Eds.). (2016). Socialbots and their friends: Digital media and the automation of sociality. Taylor & Francis.

Gentile, A., Santangelo, A., Sorce, S., & Vitabile, S. (2011). Novel human-to-human interactions from the evolution of HCI. *Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 600-605).

Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech and Language*, *25*(3), 601-634.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619-619.

Gregory, R. W., & Henfridsson, O. (2021). Bridging art and science: Phenomenon-driven theorizing. *Journal of the Association for Information Systems*, 22(6), 1509-1523.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The role of artificial intelligence and data network effects for creating user value. *Academy of Management Review*, *46*(3), 534-551.

Gregory, R. W., Kaganer, E., Henfridsson, O., & Ruch, T. J. (2018). IT consumerization and the transformation of IT governance. *MIS Quarterly*, *42*(4), 1225-1253.

Griffin, E., Ledbetter, A., & Sparks, G. G. (2018). *A first look at communication theory* (10th ed.). McGraw-Hill Education.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, *51*(1), 79-88.

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, *90*, 343-350.

Hacker, A. (2021). *"Alexa, who are you?" Analysing Alexa's, Cortana's and Siri's vocal personality.* Presented at the Konferenz Elektronische Sprachsignalverarbeitung. https://www.essv.de/paper.php?id=1095

Hall, J. A., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin*, *131*(6), 898-924.

Hartwig, M., & Bond, C. F. (2011). Why Do Lie-Catchers Fail? A Lens Model Meta-Analysis of Human Lie Judgments. *Psychological Bulletin*, *137*(4), 643-659.

Heider, F. (1920). Zur subjektivität der sinnesqualitäten [On the subjectivity of sense qualities] [Unpublished PhD dissertation]. University of Graz.

Heider, F. (1925). Ding und Medium [The thing and the medium]. *Symposium*, *1*, 109-157.

Hess, T., Fuller, M., & Campbell, D. (2009). Designing interfaces with social presence: using vividness and extraversion to create social recommendation agents. *Journal of the Association for Information Systems*, *10*(12), 889-919

Hildebrand, C., Efthymiou, F., Busquet, F., Hampton, W. H., Hoffman, D. L., & Novak, T. P. (2020). Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications. *Journal of Business Research, 121*, 364-374.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261-266.

Jennings, N., Jennings, N. R., & Wooldridge, M. J. (Eds.). (1998). *Agent technology: Foundations, applications, and markets*. Springer.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1-38

Johar, S. (2016). *Emotion, affect and personality in speech*. Springer.

Jung, H., Kim, H., & Ha, J. W. (2020). Understanding differences between heavy users and light users in difficulties with voice user interfaces. *Proceedings of the 2nd Conference on Conversational User Interfaces*.

Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed. draft). Available at https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research*, *32*(3), 713-735.

Khaled, F. (2021). *Apple removes Siri's female voice as its default and adds two new voices. Business Insider*. https://www.businessinsider.com/apple-removes-siri-female-voice-as-its-default-2021-4

Kim, H. C., Cha, M. C., & Ji, Y. G. (2021). The impact of an agent's voice in psychological counseling: Session evaluation and counselor rating. *Applied Sciences*, *11*(7), Article 2893

Klimkov, V., Ronanki, S., Rohnke, J., & Drugman, T. (2019). *Fine-grained robust prosody transfer for single-speaker neural text-to-speech.* Available at https://arxiv.org/pdf/1907.02479.pdf

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2020). Value co-creation in smart services: a functional affordances perspective on smart personal assistants. *Journal of the Association for Information Systems*, *22*(2), 418-458.

Komiak, S. X., & Benbasat, I. (2003). Understanding Customer Trust in Agent-Mediated Electronic Commerce, Web-Mediated Electronic Commerce, and Traditional Commerce. *Information Technology and Management*, *5*(1-2), 181-207.

Kovalerchuk, B. (2021). Explainable machine learning and visual knowledge discovery. http://www.cwu.edu/~borisk/pub/2020/KovalerchuK_EML_VKDca.pdf

Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking Trust in technology, humanness, and trust. *Journal of the Association for Information*, *16*(10), 880-918.

Law, J., & Rennie, R. (2015). *A dictionary of physics* (7th ed.). Oxford University Press.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Lee, I., & Shin, Y. J. (2018). Fintech: Ecosystem, business models, investment decisions, and challenges. *Business Horizons*, *61*(1), 35-46.

Lin, A. Y., Kuehl, K., Schöning, J., & Hecht, B. (2017). Understanding "death by GPS": A systematic study of catastrophic incidents associated with personal navigation technologies. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 1154-1166).

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90-103.

Lorenzo-Trueba, J., Drugman, T., Latorre, J., Merritt, T., Putrycz, B., Barra-Chicote, R., ... & Aggarwal, V. (2018). *Towards achieving robust universal neural vocoding*. Available at https://arxiv.org/abs/1811.06292

Low, C. (2020). *Alexa will seem more human with breathing pauses and learning skills*. Engadget. https://www.engadget.com/amazon-2020-alexa-breathing-teach-voice-profiles-for-kids-172918631.html

Lowry, P. B., Dinev, T., & Willison, R. (2017). Why security and privacy research lies at the centre of the information systems (IS) artefact: Proposing a bold research agenda. *European Journal of Information Systems*, *26*(6), 546-563.

Lusk, M. M., & Atkinson, R. K. (2007). Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology*, *21*, 747-764.

Malle, B. F., & Ickes, W. (2000). Fritz Heider: Philosopher and social psychologist. In *Portraits of pioneers in psychology* (pp. 227-246). Psychology Press.

Marge, M., Miranda, J., Black, A. W., & Rudnicky, A. I. (2010). Towards improving the naturalness of social conversations with dialogue systems. *Proceedings of the SIGDIAL 2010 Conference: 11th Annual Meeting of the Special Interest Group On Discourse and Dialogue* (pp. 91-94).

Markus, M. L. (2017). Datification, organizational strategy, and IS research: What's the score? *The Journal of Strategic Information Systems, 26*(3), 233-241.

Martin, J. & Finnegan, M. (2020). *What is IFTTT? How to use If This, Then That services.* Computerworld. https://www.computerworld.com/article/3239304/what-is-ifttt-how-to-use-if-this-then-that-services.html

Masina, F., Orso, V., Pluchino, P., Dainese, G., Volpato, S., Nelini, C., Mapelli, D., Spagnolli, A., & Gamberini, L. (2020). Investigating the accessibility of voice assistants with impaired users: Mixed methods study. *Journal of Medical Internet Research, 22*(9) Article e18431.

McTear, M. F. (2017). The Rise of the conversational interface: A new kid on the block? In J. Quesada, F. J. Martín Mateos, & T. López Soto, (Eds.) *Future and emerging trends in language technology* (pp. 38-49). Springer.

Melara, R. D., & Mounts, J. R. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception & Psychophysics, 56*(1), 73-90.

Metatla, O., Oldfield, A., Ahmed, T., Vafeas, A., & Miglani, S. (2019). Voice user interfaces in schools: Co-designing for inclusion with visually-impaired and sighted pupils. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.

Metcalf, L., Askay, D. A., & Rosenberg, L. B. (2019). Keeping humans in the loop: Pooling knowledge through artificial swarm intelligence to improve business decision making. *California Management Review, 61*(4), 84-109.

Miller, T. (2023). *Explainable AI is dead, long live explainable AI! Hypothesis-driven decision support.* Available at https://arxiv.org/abs/2302.12389

Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *MIS Quarterly, 45*(4), 1999-2022.

Murray, A., Rhymer, J. E. N., & Sirmon, D. G. (2021). Humans and technology: Forms of conjoined agency in organizations. *Academy of Management Review, 46*(3), 552-571.

Murtarelli, G., Gregory, A., & Romenti, S. (2021). A conversation-based perspective for shaping ethical human-machine interactions: The particular challenge of chatbots. *Journal of Business Research, 129*, 927-935.

Narynov, S., Zhumanov, Z., Gumar, A., Khassanova, M., & Omarov, B. (2021). Development of chatbot psychologist applying natural language understanding techniques. *Proceedings of the 21st International Conference on Control, Automation and Systems*.

Natale, S. (2021). *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press.

Nass, C., & Gong, L. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM, 43*(9), 36-43.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied, 7*(3), 171-181.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81-103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are Social Actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72-78).

Nederlandse Omroep Stichting. (2020). *Fout navigatiesysteem leidt auto's naar water in haven Marseille* [Faulty navigation system leads cars to water in Marseille port]. https://nos.nl/artikel/2325654-fout-navigatiesysteem-leidt-auto-s-naar-water-in-haven-marseille

Norman, D. A. (1988). Affordance, conventions, and design. *Interactions, 6*(3), 38-43.

O'Leary, D. E. (2019). GOOGLE'S Duplex: Pretending to be human. *Intelligent Systems in Accounting, Finance and Management, 26*(1), 46-53.

Pagani, M., Racat, M., & Hofacker, C. F. (2019). Adding voice to the omnichannel and how that

affects brand trust. *Journal of Interactive Marketing*, *48*, 89-105.

Pearl, C. (2016). *Designing voice user interfaces*. O'Reilly Media.

Perrakis, A., Hohenberger, W., & Horbach, T. (2013). Integrated operation systems and voice recognition in minimally invasive surgery: comparison of two systems. *Surgical Endoscopy*, *27*(2), 575-579.

Pfeffer, J., & Salancik, G. R. (1978). A resource dependence perspective. In *Intercorporate relations: The structural analysis of business.* Cambridge University Press.

Puranam, P., & Vanneste, B. S. (2022). *Artificial intelligence, trust, and perceptions of agency*. Available at https://doi.org/10.2139/ssrn.3897704

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). "Alexa is my new BFF": Social roles, user satisfaction, and personification of the Amazon Echo, *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2853-2859).

Qiu, L., & Benbasat, I. (2005). Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human-Computer Interaction*, *19*(1), 75-94.

Quesada, W., & Lautenbach, B. (2017). *Programming voice interfaces*. O'Reilly Media.

Rai, A., Constantinides, P., & Sarker, S. (2019). Next generation digital platforms: Toward human-AI hybrids. *MIS Quarterly*, *43*(1), iii-ix.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, *46*(1), 192-210.

Redeker, G. (1984). On differences between spoken and written language. *Discourse Processes*, *7*(1), 43-55.

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

Rosenthal, G. G., & Ryan, M. J. (2000). Visual and acoustic communication in non-human animals: A comparison. *Indian Academy of Sciences*, *25*(3), 285-290.

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, *26*(5), 443-452.

Ruan, S., Jiang, L., Xu, J., Tham, B. J. K., Qiu, Z., Zhu, Y., ... & Landay, J. A. (2019). Quizbot: A dialogue-based adaptive learning system for factual knowledge. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.

Rubin, D.L., Hafer, T., & Arata, K. (2000). Reading and listening to oral-based versus literate-based discourse. *Communication Education*, *49*(2), 121-133.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach.* Pearson.

Sarkadi, S., Wright, B., Masters, P., & McBurney, P. (2021). *Deceptive AI*: *First International Workshop, DeceptECAI 2020, Santiago de Compostela, Spain, August 30, 2020 and Second International Workshop, DeceptAI 2021, Montreal, Canada, August 19, 2021, Proceedings.* Springer Nature.

Schlögl, S., Chollet, G., Garschall, M., Tscheligi, M., & Legouverneur, G. (2013). Exploring voice user interfaces for seniors. *ACM International Conference Proceeding Series.*.

Schmitt, A., Zierau, N., Janson, A., & Leimeister, J. M. (2021). Voice as a contemporary frontier of interaction design. *Proceedings of the European Conference on Information Systems.*

Schöbel, S., Schmitt, A., Benner, D. Saqr, M., Janson, A., & Leimeister, J. M. (in press). Charting the evolution and future of conversational agents: A research agenda along five waves and new frontiers. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-023-10375-9

Scholz, R. W. (2017). Managing complexity: From visual perception to sustainable transitions—contributions of Brunswik's theory of probabilistic functionalism. *Environment Systems and Decisions*, *37*(4), 381-409.

Schuetz, S., & Venkatesh, V. (2020). The rise of human machines: How cognitive computing systems challenge assumptions of user-system interaction. *Journal of the Association for Information Systems, 21*(2), 460-482.

Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in human-agent interaction: A Survey. *ACM Computing Surveys*, *54*(4), 1-43.

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems, 37*(3), 875-900.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317-1323.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.

Siddarth, D., Acemoglu, D., Allen, D., Crawford, K., Evans, J., Jordan, M., & Weyl, E. (2021). *How AI fails us.* Available at https://ethics.harvard.edu/files/center-for-ethics/files/aifailsus.jhdcarr_final_2.pdf?m=1651510742

Song, S., Baba, J., Nakanishi, J., Yoshikawa, Y., & Ishiguro, H. (2020). Mind the voice!: Effect of robot voice pitch, robot voice gender, and user gender on user perception of teleoperated robots. *Proceedings of the Conference on Human Factors in Computing Systems.*

Straßmann, C., Krämer, N. C., Buschmeier, H., & Kopp, S. (2020). Age-related differences in the evaluation of a virtual health agent's appearance and embodiment in a health-related interaction: Experimental lab study. *Journal of Medical Internet Research*, *22*(4), Article e13726.

Strich, F., Mayer, A. S., & Fiedler, M. (2021). What do I do in a world of artificial intelligence? Investigating the impact of substitutive decision-making AI systems on employees' professional role identity. *Journal of the Association for Information Systems, 22*(2), 304-324.

Sujatha, J., & Rajagopalan, S. P. (2017). Performance evaluation of machine learning algorithms in the classification of Parkinson disease using voice attributes. *International Journal of Applied Engineering Research,* 12(21), 10669-10675.

Suh, K. S., Kim, H., & Suh, E. K. (2011). What if your avatar looks like you? Dual-congruity perspectives for avatar use. *MIS Quarterly, 35*(3), 711-729.

Sutton, S. J. (2020). Gender ambiguous, not genderless: Designing gender in voice user interfaces (VUIs) with sensitivity. *Proceedings of the 2nd Conference on Conversational User Interfaces*.

Sutton, S. J., Foulkes, P., Kirk, D., & Lawson, S. (2019). Voice as a design material: Sociophonetic inspired design strategies in human-computer interaction. *Conference on Human Factors in Computing Systems.*

Tamagawa, R., Watson, C. I., Kuo, I. H., Macdonald, B. A., & Broadbent, E. (2011). The effects of synthesized voice accents on user perceptions of robots. *International Journal of Social Robotics, 3*(3), 253-262.

Tarantola, A. (2020, November 11). *Alexa is getting better at guessing your intentions*. Engadget. https://www.engadget.com/alexa-is-getting-better-at-guessing-your-intentions-202813759.html

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human-robot interaction. *Computers in Human Behavior, 38*, 75-84.

Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J., Bernstein, A., & Leimeister, M. (2021). Female by Default? Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. *Proceedings of the Conference on Human Factors in Computing Systems.*

Torre, I., Goslin, J., & White, L. (2020). If your device could smile: People trust happy-sounding artificial agents more. *Computers in Human Behavior*, *105*, Article 106215.

Treasure, J. (2020, August 3). The 4 ways sound affects us. *JT.* https://www.juliantreasure.com/blog/4-ways-sound-affects

Trovato, G., Lopez, A., Paredes, R., & Cuellar, F. (2017). Security and guidance: Two roles for a humanoid robot in an interaction experiment. *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 230-235).

Truschin, S., Schermann, M., Goswami, S., & Krcmar, H. (2014). Designing interfaces for multiple-goal environments: Experimental insights from in-vehicle speech interfaces. *ACM Transactions on Computer-Human Interaction, 21*(1), Article 7.

Vance, A., Lowry, P. B., & Eggett, D. (2015). Increasing accountability through user-interface design artifacts. *Source: MIS Quarterly*, *39*(2), 345-366.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention Is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems.*

Villazon, L. (2021). *What is the time resolution of our senses?* Science Focus. https://www.sciencefocus.com/the-human-body/what-is-the-time-resolution-of-our-senses/

Vimalkumar, M., Sharma, S. K., Singh, J. B., & Dwivedi, Y. K. (2021). "Okay Google, what about my privacy?" User's privacy perceptions and acceptance of voice based digital assistants. *Computers in Human Behavior, 120*, Article 106763.

Völkel, S. T., Schödel, R., Buschek, D., Stachl, C., Winterhalter, V., Bühner, M., & Hussmann, H. (2020). Developing a personality model for speech-based conversational agents using the psycholexical approach. *Proceedings of the Conference on Human Factors in Computing Systems*.

Wang, W., & Benbasat, I. (2016). Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents. *Journal of Management Information Systems, 33*(3), 744-775.

Weber, R. (2012). Evaluating and developing theories in the information systems discipline. *Journal of the Association for Information Systems*, *13*(1), 1-30.

Wood, R. E. (1986). Task complexity: Definition of the construct. Organizational behavior and human decision processes, *37*(1), 60-82.

World Health Organization. (2019). *World report on vision*. Available at https://www.who.int/publications/i/item/9789241516570

Zepf, S., Gupta, A., Krämer, J. P., & Minker, W. (2020). EmpathicSDS: Investigating lexical and acoustic mimicry to improve perceived empathy in speech dialogue systems. *ACM International Conference Proceeding Series*.

Zhang, P., Benbasat, I., Carey, J., Davis, F., Galletta, D. F., & Strong, D. (2002). Human-computer interaction research in the MIS discipline. *Communications of the Association for Information Systems, 9*(20), 334-355.

Zhang, P., & Li, N. (2005). The intellectual development of human-computer interaction research: A critical assessment of the MIS literature (1990-2002). *Journal of the Association for Information Systems, 6*(11), 227-292.

Zhang, Y., Liao, Q. V., & Bellamy, R. K. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 295-305).

Zhao, X., & Malle, B. F. (2022). Spontaneous perspective taking toward robots: The unique impact of humanlike appearance. *Cognition, 224*, Article 105076.

Zierau, N., Hildebrand, C., Bergner, A., Busquet, F., Schmitt, A., Leimeister, J. M. (2022). Voice bots on the frontline: Voice-based interfaces enhance flow-like consumer experiences & boost service outcomes, *Journal of the Academy of Marketing Science*, *51*(4), 823-842

Appendix

For some of our constructs, no exact assessment exists. Constructs can be measured or construct specification can begin by using or adapting self-reported and independent measures from previous studies.

Table A1 Assessment of the Conceptual Model's Key Constructs

Construct	Definition	Assessment	
		Self-reported	Independent
Agency attribution	In a specific instance, a user's perceptual estimation of an artifact's agency based on own beliefs and experience with an artifact.	Perceived agency (Jennings et al., 1998; Puranam & Vanneste, 2022)	/
NLP performance	The extent of an artifact's ability to understand and respond to human language as it is spoken and written.	/	Performance metrics including accuracy, precision, recall, or word error rate (Billsus & Pazzani, 1999; Sujatha & Rajagopalan, 2017)
Naturalness of voice	Replicability of human auditory cues in the design of an artifact.	Humanness of artifact's voice (Lankton et al., 2015; Schuetzler et al., 2020)	Use of natural (versus robotic) voice (Cowan et al., 2006)
Degree of autonomy	A dimension of agency referring to an artifact's ability to perform (parts of) a task independently of direct human intervention by being in control of own actions and state, and by leveraging previous interactions.	Artifact's ability to control own actions and leverage prior information gathered (Gray et al., 2007; Jennings et al., 1998)	Degree of decision-making latitude (Baird & Maruping, 2021)
Degree of flexibility	A dimension of agency referring to an artifact's ability to to adapt to an environment in a proactive and social manner.	Artifact's goal-directedness, responsiveness, sociability, and self-control (Gray et al., 2007; Jennings et al., 1998; Schuetzler et al., 2020)	Artifact's computational and interfacing capabilities (Baird & Maruping, 2021)
Degree of situatedness	A dimension of agency referring to an artifact's ability to receive, react to, and influence sensors in an environment.	Artifact's contextuality (Gray et al., 2007; Jennings et al., 1998)	Artifact's awareness capabilities (Baird & Maruping, 2021)
System uncertainty	The degree to which an artifact's true attributes and nature cannot be accurately anticipated or predicted due to imperfect information.	Uncertainty, trustworthiness, or transparency of artifact (Pavlou et al., 2007)	Use (presence) of explanatory methods and analyses; presence of system disclosure (Jacovi et al., 2021; Wang & Benbasat, 2016)
Context uncertainy	The degree to which the (future) states of the environment cannot be accurately anticipated or predicted due to imperfect information.	Uncertainty or felt risk (Pavlou et al., 2007; McKnight et al., 2022)	Information asymetry and the environmental dimensions of a task environment (Dess & Beard, 1984)
User familiarity	A dimension of user characteristics referring to a user's previous experiences with, interactions with, and learning of similar artifacts.	Algorithmic familiarity (Gefen, 2000; Komiak & Benbasat, 2006)	Historical user data and time spent with system
User similarity	A dimension of user characteristics referring to the perceived similarity between an artifact's voice and the user's voice in terms of auditory cues.	/	User matching, i.e., regarding the pitch, gender or accent of a voice (Lubold et al., 2020; McGinn & Torre, 2019; Suh et al., 2011)

About the Authors

Anuschka Schmitt is a research associate and PhD candidate at the Institute of Information Management at the University of St.Gallen (HSG), Switzerland and a visiting research fellow at Harvard University's School of Engineering and Applied Sciences. Anuschka is passionate about the societal impact of AI systems, focusing on the role that conversational AI has on human perception and behavior. Her research mainly engages in user trust and decision-making in knowledge work and educational domains.

Naim Zierau is a postdoctoral researcher at the University of St.Gallen. His research focuses on how AI-based information systems and their design affect user experience and business outcomes. He is also concerned with the organizational implications of deploying AI technologies.

Andreas Janson is a postdoctoral researcher at the Institute of Information Management (IWI-HSG) at the University of St.Gallen, Switzerland. His dissertation investigates how to design digital learning processes. His research interests focus on issues relating to the design of digital services and platform ecosystems. His research has been published in leading information systems and management journals such as the *Journal of the Association for Information Systems*, *European Journal of Information Systems*, *Journal of Information Technology*, and *Academy of Management Learning and Education*, and in the proceedings of the Hawaii International Conference on System Sciences, the European Conference on Information Systems, and the International Conference on Information Systems. He has also received Best Paper nominations at major conferences and received a Best Paper award at the Hawaii International Conference on System Sciences 2020 as well as the Vinton G. Cerf Award at DESRIST 2020.

Jan Marco Leimeister is the chaired professor and managing director of the Institute of Information Management at the University of St.Gallen, Switzerland. He is also a director at the Research Center for IS Design at the University of Kassel, Germany. His work covers digital business, digital transformation, digital service management, crowdsourcing, digital work, digital learning services and collaboration engineering. Jan Marco has been internationally recognized for outstanding research, teaching, and education. Since 2009, he has repeatedly ranked among the top 1% of the most productive researchers and professors in business administration in the German-speaking area. He serves on the AIS leadership council, is co-editor-in-chief of the *Journal of Information Technology* and serves on the editorial board of *Journal of Management Information Systems* and *Information Systems Research*. He is also an entrepreneur and serves as a senior advisor, board member, and keynote speaker for national and international organizations.