Towards a Better Understanding of Information Disclosure in Human-AI Interactions

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

M.Sc. Miriam Elisabeth Gieselmann

aus Bielefeld

Tübingen

2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	29.01.2024
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Kai Sassenberg
2. Berichterstatter/-in:	Prof. Dr. Sonja Utz

Content

Tables and Figures

Tables

Figures

Chapter 1: General Introduction

We interact with AI (Artificial Intelligence) technologies[1] on a regular basis. Applications include (but are not limited to) conversational AI assisting users in private contexts (e.g., Amazon Alexa or Google Assistant), algorithms recommending what to purchase, watch, and eat (e.g., in online shopping, streaming services, or recipe recommendations), or AI systems supporting crucial decisions in business contexts (e.g., hiring or investment decisions). And still, technological progress allows for the development of evermore complex and capable AI systems to assist, augment, or even replace humans in different tasks (e.g., Hassani et al., 2020). For example, the recent introduction of Chat-GPT (Open AI, 2023) was a milestone in technology development (Nayyar, 2023), surpassing longstanding expectations about potential AI capabilities. After five days, Chat-GPT had already acquired one million users (Duarte, 2023) and this number increased to more than 100 million by February 2023 (i.e., two months after the launch of Chat-GPT; Milmo, 2023) – showing that the introduction of such easily accessible and highly capable tools can lead to an unprecedented and rapid mass adoption of AI.

The current developments in the field of AI are, on the one hand, associated with many hopes, such as enhancing comfort and convenience, offering better-personalized services, increasing productivity and efficiency, reducing costs, or enabling more objective and data-driven decisions (e.g., Brynjolfsson & Mitchell, 2017; Hang & Chen, 2022; Sundar, 2020). On the other hand, the widespread application of AI also comes with severe risks and concerns – such as moral and ethical dilemmas, potential biases and discrimination (e.g., Hong et al., 2020), or threats to data security and privacy (e.g., Yan et al., 2022). Given these potentials and risks as well as the increasing adoption of AI technologies, a deeper understanding of the psychological aspects of human-AI interactions is crucial (cf. Sundar, 2020).

In cases where AI is designed to produce personalized output, such as conversational AI or recommender systems, privacy concerns are a particularly relevant issue. These systems typically require large amounts of personal data (i.e., detailed information about the individual user) to optimize their results and, furthermore, come with an ever-increasing capacity to collect, store, and process these data. The vulnerability of privacy became evident, for instance,

[1] There is an ongoing debate about the exact definition of AI (cf. Kok et al., 2009). In this dissertation, I refer to AI as any kind of 'automation of intelligent behavior' (Luger, 2009, p. 1). More specifically, AI describes a technology's ability 'to perform tasks commonly associated with intelligent beings … such as the ability to reason, discover meaning, generalize, or learn from past experience' (Copeland, 2023). Following this working definition, the term AI describes a broad range of technologies whose actions can be interpreted as intelligent – thus laying the focus on the technologies' capabilities as perceived and interpreted by the human instead of details of the technical implementation.

through several privacy breaches that occurred during the last years, such as the Cambridge Analytica Scandal (cf. Confessore, 2018) or several scandals about privacy violations in connection with conversational AI (cf. Bleich, 2018; Lynskey, 2019). Despite being aware of potential privacy risks, users still frequently share detailed personal information with such systems (Pal et al., 2020). Given this and the ongoing proliferation of AI systems, it is of particular interest to gain further insights into the determinants of personal information disclosure in the interaction with these technologies.

Thus, the current dissertation strives toward a better understanding of humans' disclosure towards AI technologies. Building on the ideas of the Computers Are Social Actors Paradigm (CASA; Reeves & Nass, 1996) and anthropomorphism (Epley et al., 2007), the dissertation takes a human-oriented approach – considering humans' perception of the technology rather than the technical details of implementation. It focuses on the role of two potential determinants of disclosure, namely specific perceptions of (1) the *interaction partner* (i.e., the technology and the provider) and (2) the *interaction* (i.e., perceived interactivity and output quality) from the perspective of people directly interacting with AI technology. Going beyond this user-centered perspective (where decisions mainly impact the individual user), this dissertation further takes into account that, outside of private use contexts of AI, the choices made by few decision-makers often have consequences for large groups of people – for instance, when managers decide about implementing AI in work-related contexts. Given the large impact of their decisions, the perspective of decision-makers is of particular relevance. Thus, this thesis further addresses how characteristics of the interaction partner (i.e., the technology) and the necessity to disclose personal data impact decision-makers' AI acceptance.

Privacy, Disclosure, and Personalization

In order to gain a better understanding of disclosure in human-AI interactions, it is, in the first step, necessary to get an idea of the concept of privacy – and its particular relevance for disclosure in interactions with AI that produces personalized outputs. In general, privacy can refer to different things such as the non-intrusion of one's physical space (i.e., physical privacy), the non-interference involving one's choices (i.e., decisional privacy), or the non-intrusion of one's thoughts and personal identity (i.e., mental privacy; Tavani, 2008). The focus of this dissertation, however, is *informational* privacy, which refers to the extent to which an individual has control over their personal information or the extent to which their personal data is protected (Tavani, 2008). A lack of informational privacy is not only a threat to individual autonomy (as the individual no longer has control over their data) but can, for instance, also involve the danger of constant surveillance (e.g., by authoritarian regimes; cf. Westin, 2003),

discrimination (e.g., when personal data such as race, sexual orientation, political or religious beliefs, is revealed without the individual's consent; cf. Westin, 2003), or cybercrimes (e.g., identity theft or fraud; cf. Milne et al., 2004). Hence, this kind of privacy is seen as a fundamental good that needs to be secured, as becomes evident also in various legal efforts to protect it (e.g., GDPR; European Union, 2018). However, it is nowadays severely threatened by technologies that collect vast amounts of personal data (Tavani, 2008).

As described before, in order to make data-driven decisions, AI needs large amounts of data. In many cases, it is not enough that the AI has an extensive training data set to begin with, but many AI systems (e.g., those based on machine learning techniques) are designed to incorporate new data in order to improve their performance continuously (Glikson & Woolley, 2020). While all interactions with AI technology can potentially threaten informational privacy as soon as personal data is shared, this threat is especially apparent in cases where AI is designed to personalize output – such as giving fitting recommendations or identifying the content most relevant for the user. Here, the AI needs detailed personal information linked to the individual user to obtain a high-quality personalized output. Thus, private companies, which are typically the providers of such AI systems, regularly gather, process, store, and analyze vast amounts of personal data to gain competitive advantages (Dinev & Hart, 2006).

Hence, personalization is well-known as a double-edged sword: while highly personalized output is often perceived as beneficial, it is also known to evoke privacy concerns (Chellappa & Sin, 2005; Montgomery & Smith, 2009) – especially for smart technologies (Hepp et al., 2022). Users might, for instance, worry whether their data is appropriately protected from unauthorized access (e.g., being hacked or breached), sold or shared with third parties (e.g., marketing firms), or used for other than the declared purposes (Dinev & Hart, 2006).

Despite these concerns, people still routinely interact with AI technologies that require disclosing large amounts of personal data (such as conversational AI; Kinsella, 2023). While a large part of data collection happens through implicit tracking (e.g., by monitoring transactions), people also explicitly disclose personal information to AI technologies (Zimmer et al., 2010) or at least consciously agree to their data being collected. This behavior is in line with earlier observations described as the *privacy paradox* – referring to people disclosing personal information despite indicating to be concerned about privacy (for a review see Kokolakis, 2017). On the one hand, this idea has received empirical support in numerous studies in the context of e-commerce (e.g., Beresford et al., 2012; Spiekermann et al., 2001) and social media (e.g., Reynolds et al., 2011; Taddicken, 2014; Tufekci, 2008). On the other hand, there are also

several studies that question the relevance and existence of the privacy paradox. These studies propose that people actually do care for their privacy (e.g., uninstall apps that collect data which they do not want to share; Boyles et al., 2012; or prefer to rule out secondary use of their data over a 10% price cut; D'Souza & Phelps, 2009) and find other ways than non-disclosure to protect it (e.g., on social media; Miltgen & Peyrat-Guillard, 2014). Beyond that, a more recent meta-analysis showed that the relationship between privacy concerns and information disclosure is small but, nonetheless, significant (Baruh et al., 2017).

Addressing these somewhat inconsistent findings, it has been proposed that people's decision to disclose despite being concerned might not be paradoxical but rather based on a rationale *privacy calculus* – indicating that people weigh up costs and benefits in disclosure decisions (e.g., Dinev & Hart, 2006; Krasnova et al., 2010) and disclose more when the perceived benefits exceed their concerns (e.g., Kezer et al., 2022). Developed initially to understand e-commerce transactions, this idea can be easily transferred to other online interactions as well as human-AI interactions in general: to get desirable outcomes – such as access to restricted content, social connection, convenience, or personalized recommendations – people often willingly choose to give up certain degrees of privacy.

Adding to the idea of the privacy calculus – a very general approach considering perceived risks and perceived benefits as determinants of disclosure – this dissertation will consider more specific determinants by taking a human-oriented approach that will be explained and elaborated in the following section. Further, as the context of a disclosure decision plays a crucial role (Bol et al., 2018), this dissertation adds to the literature that has (so far) predominantly focused on disclosure in e-commerce and on social media (for a review, see Kokolakis, 2017) by considering the increasingly common context of human-AI interactions.

Investigating Human-AI Interactions: A Human-Oriented Approach

According to the Computers Are Social Actors (CASA) paradigm, technologies are often treated as social actors (Nass & Moon, 2000; Reeves & Nass, 1996). CASA implies that people mindlessly apply (social) scripts (i.e., mental representations that affect perception, comprehension, and behavior in specific events; Schank & Abelson, 1977) from human-human interactions to human-technology interactions (Nass & Moon, 2000). Hence, humans are expected to use the same schemas and rules in human-technology interaction as in human-human interactions (Reeves & Nass, 1996). More specifically, when technologies depict sufficient social cues, humans are considered to often treat these technologies socially instead of investing cognitive effort to determine another response that is possibly more suitable to the technology (Reeves & Nass, 1996). As a result, humans mindlessly ascribe technologies human-like

characteristics (e.g., personality traits), apply stereotypes and norms, and exhibit social behaviors – despite understanding that the technology is not human (Nass & Moon, 2000; Reeves & Nass, 1996).

Adding to the ideas of CASA, the concept of anthropomorphism describes humans' tendency to ascribe humanlike characteristics, mental states, motivations, intentions, and emotions to non-human entities (Epley et al., 2007). Again, humans do not have to believe that this entity actually possesses the ascribed characteristics but can still act *as if* it did (Epley et al., 2007). However, anthropomorphism does not require mindlessness but can be both mindless (as proposed by CASA) or mindful (Kim & Sundar, 2012), the latter being assumed to occur more likely the more social cues a technology offers (Lombard & Xu, 2021).

Since the introduction of CASA and anthropomorphism, people, technologies, and their interactions have changed (cf. Gambino et al., 2020): nowadays, people are used to having more frequent, complex, and variant interactions with increasingly complex and capable technology. Nonetheless, these ideas continue to inform our understanding of how people interact with technologies (cf. Gambino et al., 2020). While humans' experience with technology interactions might lead to the development of specialized scripts for such interactions (Gambino et al., 2020), the rapid progress in AI technologies still requires people to interact with evermore complex technologies for which they (currently) do not have a fitting script – potentially evoking the application of existing scripts from human-human interaction.

Furthermore, modern (AI) technologies are often intentionally designed to facilitate the application of anthropocentric knowledge – a relevant determinant of anthropomorphism (Epley et al., 2007) – by using human-like design cues (e.g., voice outputs, language style, movement patterns, or facial characteristics). Besides, they often possess rich social cues such as being interactive, personalizing the responses to their interaction partner, and taking over roles that were traditionally filled by humans (Nass & Moon, 2000). As a result, people can often no longer reliably distinguish whether they are encountering (the output of) another human or an AI technology (e.g., Köbis & Mossink, 2021; Warwick & Shah, 2016), which further increases the likelihood of applying knowledge and scripts from human-human interactions.

However, even in cases where humans are aware of the AI's non-human nature, anthropomorphism provides a strategy to predict, understand, and explain the actions of AI (cf. Epley et al., 2007) – which is often perceived as a black box with intransparent and complex decision-making processes (Castelvecchi, 2016; Glikson & Woolley, 2020) – by imbuing human reasoning. While the conclusions people draw from anthropomorphizing technologies

might not always be accurate, anthropomorphism can still facilitate human-technology interaction (Epley et al., 2007).

Based on the previously described ideas, it is highly plausible to draw on insights from human-human interactions in order to understand human-AI interactions, including the disclosure of personal information. Following this human-oriented approach, the current dissertation investigates whether determinants of disclosure in human-human interactions are also relevant for disclosure in human-AI interactions. As will be elaborated in the following sections, the role of perceptions of (1) the *interaction partner* and (2) the *interaction* as determinants of disclosure will be addressed.

Perceptions of the Interaction Partner as Determinants of Disclosure

Perceptions of an interaction partner might be crucial for the decision whether to disclose personal information. According to social evaluation theories (for an integrated approach, see Abele et al., 2021), people perceive other humans along two core dimensions: a task-related (agency, competence, or "getting ahead") and a social-emotional dimension (communion, warmth, or "getting along"). Showing significant overlap with these ideas, humans are assumed to rate others' trustworthiness depending on perceptions of competence, benevolence, and integrity (Mayer et al., 1995). In a nutshell, these dimensions can be summarized as whether the interaction partner is perceived as capable of implementing their intentions and whether these intentions are good (cf. Abele et al., 2021).

These task-related and socio-emotional perceptions of another person shape interaction behavior in human-human interactions (e.g., cooperation; Balliet & van Lange, 2013; help, neglect, and harassment; Cuddy et al., 2007, 2008; or negotiation behavior; Swan et al., 1999). Further, they are of particular relevance when deciding to make oneself vulnerable (cf. Mayer et al., 1995) – for instance, by disclosing personal information to an interaction partner, who could potentially misuse it for other than the intended purpose. This notion is supported by empirical evidence showing, for example, that benevolence (as a social-emotional perception of the interaction partner) is related to disclosure-based trust (e.g., discussing problems that could be used against oneself; Qiu et al., 2022), that trustworthiness predicts disclosure of highly sensitive information (Boon & Miller, 1999), and that people disclose more information to those whom they initially like (for a meta-analysis, see Collins & Miller, 1994).

To sum up, perceptions of another human as an interaction partner drive interaction behavior in human-human interactions, including the disclosure of personal information. This is also the case when humans communicate via technology (i.e., computer-mediated communication, see Excursion 1). The focus of this dissertation is, however, on interaction *with* technology instead of interaction *via* technology. While in human-human interaction it is evident that one communicates with another human, the interaction partner is less clearly defined in human-technology interactions. Here, at least two potential interaction partners can be considered: the technology itself and the provider 'behind' the technology (cf. Tschopp et al., 2021). As will be pointed out in the following sections, perceptions of these different interaction partners are likely to shape human-AI interactions similarly to perceptions of the other human in human-human interactions.

Excursion 1: Communication via Technology

In many cases, people use technology as a medium to communicate with another human being. An extensive stream of research focuses on this so-called computer-mediated communication including interactions with other humans via mobile phones, e-mail, social media, or even robots (cf. Yao & Ling, 2020). While perceptions of the other human shape behavior also in computer-mediated interactions (e.g., T. Kim & Read, 2021), also the communication medium (i.e., the technology) can alter these perceptions. For example, people evaluate their interaction partner more positively when resolving conflicts in a video chat compared to a face-to-face interaction (Shin et al., 2017). Thus, when communicating with other humans via technology, both perceptions of the other human as interaction partner as well as the technology determine interaction behavior.

The Technology as Interaction Partner

Nowadays, communication is no longer limited to human-human conversations but can also happen in the interaction *with* technology (Guzman, 2018). As many scholars have pointed out, machines are perceived to become more and more agentic (e.g., Glikson & Woolley, 2020; Sundar, 2020) and can actively contribute to communications with the user (Sundar, 2020) – thus, no longer being a 'mere' technology or communication medium but also a potential interaction partner itself (Bankins & Formosa, 2020; Guzman, 2018). While modern technologies (e.g., digital assistants such as Siri) are often designed to be both a communication channel (e.g., when initiating phone calls to other humans) as well as a distinct source of

communication (i.e., when engaging in dyadic exchange; Guzman, 2018; Kim & Sundar, 2012), this dissertation focuses on the latter case.

Supporting the idea of technologies as distinct interaction partners, and going beyond merely treating technologies as social actors (CASA; Nass & Moon, 2000; Reeves & Nass, 1996) or anthropomorphizing (Epley et al., 2007) them, it has been observed that people ascribe mind[2] in terms of agency (e.g., self-control, planning, communication, thought) to technical systems (H. M. Gray et al., 2007). Agency constitutes the first dimension of mind (H. M. Gray et al., 2007) and relates to the afore-described task-related dimension of person perception. Since the publication of Gray and colleagues' influential paper on mind perception in 2007, technologies have developed rapidly. As modern AI technologies often have control over important resources (such as detailed information on the user's interests), develop new capacities, adapt to changes, process diverse information, and act unpredictably, ascribing agency (i.e., task-related competencies) becomes even more likely (Shank et al., 2019). However, technologies are typically perceived to lack experience (e.g., fear, pain, desire, personality), which constitutes the second dimension of mind (H. M. Gray et al., 2007) and relates more strongly to the socio-emotional dimension of person perception. While research has focused on antecedents of perceiving both cognitive (i.e., task-related) and emotional characteristics in AI (as facets of trust in AI; cf. Glikson & Woolley, 2020), and anthropomorphic as well as social design of technologies are becoming increasingly common, empirical research shows that technologies depicting experience (i.e., the capacity to feel) are still perceived as more uncanny than technologies showing agency (Appel et al., 2020). Based on these insights, it seems that people are more willing to perceive technologies as agentic interaction partners than to ascribe them to the capacity to feel. Furthermore, it can be concluded that perceptions of experience (related to the social-emotional dimension of person perception) in technologies do – at least currently – not evoke positive reactions (e.g., trusting behaviors such as disclosure) but rather negative ones. On the contrary, the role of the more task-related dimension (related to agency) is less clear, as will be elaborated in the following paragraphs. Thus, the present dissertation focuses on task-related rather than socio-emotional perceptions of AI characteristics as determinants of human interaction behavior.

[2] As pointed out by Shank et al. (2019), the question of whether humans perceive AI technologies to have a mind and which consequences this might have is distinct from the question of whether AI technologies can or will ever actually have a mind – which is a highly debated topic in cognitive science and philosophy and goes beyond the scope of this argumentation.

As technology is seen as a distinct interaction partner, it is not surprising that perceptions of its characteristics determine how humans behave when encountering them (cf. Epley et al., 2007). Focusing on task-related perceptions of the technology, it has been observed that functionality (as a technical characteristic, see Excursion 2) is related to technology acceptance, usage, and adoption (e.g., Dehghani, 2018; McLean & Osei-Frimpong, 2019; Moussawi et al., 2022). Moreover, considering a more human-oriented approach, associations between competence perceptions and, for instance, willingness to use and believability have been found (e.g., Demeure et al., 2011; X. S. Liu et al., 2022; Pitardi & Marriott, 2021).

Excursion 2: Investigating Human-AI Interaction With a Focus on Technical Features

In addition to the human-oriented approach used in the current dissertation, humans' interactions with technology are often investigated with a stronger focus on the technical features of the technologies in question. The central goal of this Human-Computer Interaction (HCI) research is to optimize and facilitate the interaction between humans and technologies (Guzman, 2018; Vollrath, 2017). Thus, it primarily focuses on the design of human-machine interfaces in order to optimize usability and user experience (Vollrath, 2017). Well-known models in HCI research include the Technology Acceptance Model (TAM; Davis et al., 1989) and the Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh et al., 2003).

There are also first clues that perceptions of task-related characteristics of technologies might impact disclosing behavior in particular. Perceived benefits – for example, increased personalization or usefulness – are connected to a higher willingness to disclose personal information (Chellappa & Sin, 2005; Sharma & Crossler, 2014; Xu et al., 2013). Accordingly, technologies perceived as capable of offering such benefits might enhance the willingness to disclose. Nonetheless, the impact of perceived task-related characteristics of technologies as interaction partners on disclosing behavior is not yet well-understood and is, thus, addressed in the current dissertation (in Chapter 2).

The Provider as Interaction Partner

As technologies are perceived to become more and more agentic, the providers of these technologies become less salient as interaction partners. While the provider of a shopping website is typically very prominently visible in the interaction (e.g., through a prominently placed brand logo), providers of technologies perceived as more agentic are often less dominant in the interaction experience – but no less relevant. For example, when a user shares personal data

with a conversational AI or a recommendation algorithm, they not only disclose this information to the technology but also entrust the technology provider to treat and secure their data appropriately. Thus, technology providers can be regarded as (one of several) interaction partners in any human-AI interaction.

Accordingly, as for the afore-described interaction partners, perceptions of the technology providers might also shape how humans interact with (AI) technologies. However, it has to be considered that technology providers differ from human interaction partners and technology as interaction partner in at least two crucial characteristics. First, the interaction with the provider is less salient and less direct than with the other types of interaction partners. Second, most technologies are provided by large, private companies with a strong profit orientation (as is the case, for example, for the most popular conversational AI's Amazon Alexa, Google Assistant, or Apple's Siri). These companies collect large amounts of personal data from many individuals, adding up to a considerable monetary value (cf. Mediavision Interactive, 2021), and there is little guarantee for the users that the companies will not act in an undesirable, unethical, or opportunistic manner (Zimmer et al., 2010).

Nonetheless, provider trustworthiness is also characterized by trusting beliefs about competence, benevolence, and integrity (McKnight et al., 2002; McKnight & Chervany, 2001) and plays a central role in people's interaction behavior – as is the case for human interaction partners (Mayer et al., 1995). More specifically, provider trustworthiness determines whether people are willing to use technologies: it is associated, for instance, with the adoption of medical engineering products (Nienaber & Schewe, 2014), the decision to purchase products online (McKnight & Chervany, 2001), the use of document management technologies (J.-H. Lee & Song, 2013) or social networking websites (Fogel & Nehmad, 2009), and the usage continuance intentions of AI technologies (Pal et al., 2020).

As provider companies naturally have a strong interest in gathering as much information about the technology users as possible to get competitive advantages (Zimmer et al., 2010), perceptions regarding their benevolence and integrity (i.e., rather socio-emotional characteristics) should be especially relevant for disclosure decisions. As pointed out, sharing information with a technical system always involves entrusting this information to the provider. Therefore, it is not surprising that perceptions of provider trustworthiness seem to be related to whether people (intend to) disclose personal information online (Joinson et al., 2010; Krasnova et al., 2010; Malhotra et al., 2004; Mesch, 2012; Taddei & Contena, 2013; Zimmer et al., 2010). The current dissertation aims to further add to the understanding of how provider trustworthiness (particularly, regarding socio-emotional aspects) impacts disclosure in human-AI

interactions (addressed in Chapter 4), focusing also on the interplay with perceived characteristics of the interaction (as another potential determinant of disclosure addressed in the following section).

In summary, the previous sections argued that perceptions of the interaction partner are relevant drivers of behavior in human-AI interaction. To better understand the impact of perceived characteristics of the interaction partner on disclosure, the current dissertation differentiates between (1) the *technology* (addressed in Chapter 2) and (2) the *provider* (addressed in Chapter 4) as relevant interaction partners.

Perceptions of the Interaction as Determinants of Disclosure

In addition to the characteristics of the interaction partner, other aspects of an interaction can also be relevant for people's behavior. As a matter of fact, each interaction can be described as a *process* that results in an *outcome* – as it has been done even in the earliest communication theories. Already Aristoteles postulated that communication between humans is a process where a speaker directs a speech toward an audience, which will result in an effect (i.e., an outcome; Krapinger, 2018). Since then, several more fine-grained communication theories, often focusing on specific aspects, have been proposed – such as the well-known communication theory by Shannon and Weaver (1964), focusing on the interaction as a *process*, or the predicted outcome value theory by Sunnafrank (1986) focusing on the role of (expected) interaction *outcome*s. Following the core ideas of these two theories, the current dissertation aims to address perceptions of specific characteristics of (1) the interaction process and (2) the interaction outcome as potential drivers of interaction behavior.

The Interaction Process

Shannon and Weaver's communication theory focuses on interaction as a process where information is transmitted via a communication channel from one interaction partner to the other one (Shannon & Weaver, 1964). Their model highlights that this process can be disturbed by different kinds of noises, such as technical problems in information transmission or background noises (Shannon & Weaver, 1964). Beyond these examples, obviously, other factors can also disturb the interaction process and, thus, act as noise in the way described by Shannon and Weaver (1964). For example, some communication channels might generally impede interaction processes more than others as they offer the human communicator only limited control over the interaction process and the exchanged information (e.g., potential information access by other than the intended interaction partner), lack possibilities to create a reciprocal communication (e.g., allowing only one-sided communication), or provide only a low speed of

information transmission (e.g., require long times for information transfer). According to Y. Liu (2003), these three factors constitute the three facets of perceived *interactivity*: (1) active control, (2) reciprocal interaction, and (3) synchronicity.

Perceived interactivity originally refers to the feeling of responsiveness of another human. However, it can, more generally, also be described as perceiving responsiveness from another entity which can also be a technology (Lew et al., 2018). As such, interactivity is not only a key factor in the interaction process between two human communicators but has also become a core evaluation criterion for technologies. As will be described in more detail in Chapter 3 of this dissertation, empirical findings propose that perceived interactivity might not only shape interaction behavior and disclosure in human-human interactions (e.g., Green et al., 2016; X. Liu et al., 2020; McLaughlin & Cody, 1982; Walsh et al., 2020) but also in interactions between humans and technologies (e.g., Martelaro et al., 2016; Y. W. Park & Lee, 2019; Złotowski et al., 2017). However, previous research has predominantly focused on one of the named interactivity facets, whereas studies investigating the relative importance of the interactivity facets for disclosure are scarce. Thus, the current dissertation takes a closer look at the different facets of perceived interactivity as a crucial characteristic of the interaction process to understand human interaction behavior regarding the disclosure towards AI technologies (in Chapter 3).

The Interaction Outcome

Once the interaction process is completed, there is, in any case, a more or less favorable outcome. As proposed by the predicted outcome value theory (Sunnafrank, 1986), an interaction's (expected) outcome plays a crucial role in human-human interactions. More specifically, the theory assumes that the expectation of sufficiently positive outcomes is a primary driver of interaction behavior. In line with the theory, positive relationships have been observed between predicted positive (relational) outcomes and interaction behavior (e.g., the amount of verbal communication and non-verbal affiliative expressiveness; Ramirez et al., 2010; Sunnafrank, 1988). Moreover, outcome values of initial interactions determine outcome expectations in subsequent interactions, which in turn influence future interaction behavior (Marek et al., 2004; Sunnafrank & Ramirez, 2004). Accordingly, actual interaction outcomes, as a central determinant of predicted outcome values in future interactions with the same interaction partner, are an important driver of interaction behavior. This seems to apply also for disclosure decisions in human-human interactions, as positive associations between perceived outcome value and communication intimacy have been reported as well (Ramirez et al., 2010; Sunnafrank, 1988).

Interaction outcomes are also crucial in technology interactions: people are more likely to use technologies when they perceive the (potential) interaction outcomes as beneficial – for instance, in terms of usefulness (cf. Davis et al., 1989) or performance expectancy (cf. Venkatesh et al., 2003). In line with the idea of the privacy calculus (Dinev & Hart, 2006), briefly introduced at the beginning of this dissertation, interaction outcomes should also be crucial for disclosure, as beneficial outcomes might outweigh the costs of disclosing personal information. Supporting this notion, it has been observed that people are willing to share personal information, for example, in exchange for beneficial outcomes such as monetary benefits (Beresford et al., 2012; Carrascal et al., 2013; Yang & Wang, 2009), personalization (Chellappa & Sin, 2005), enjoyment, or usefulness (Sharma & Crossler, 2014) – the latter displaying typical positive outcomes of technology interactions.

To sum up, perceived output quality as a core characteristic of the outcome in interactions with technologies could be a relevant determinant of interaction behavior in general and disclosure of personal information in specific. Thus, it will be considered as another potential determinant of disclosure (in Chapter 4), besides perceived interactivity as a characteristic of the interaction process. Comprising the previous sections, this dissertation aims at investigating how perceptions of the *interaction partner* (referring to the technology as well as the provider) and the *interaction* (in terms of perceived interactivity in the interaction process and perceived output quality) impact disclosure as will be described in the following section.

The Current Dissertation

Disclosure of personal information is crucial for many AI applications to work properly and, at the same time, is one of the most debated risks of AI usage (i.e., in terms of privacy concerns and data exploitation). With the widespread adoption of AI as well as the increasing possibilities to collect, store, and process large amounts of data collected while using these systems, it is evermore necessary to grasp a good understanding of when people are willing to share personal information with such technologies. Thus, the current dissertation aims at adding to the understanding of people's disclosure towards AI by investigating the impact of specific perceived characteristics of (1) the *interaction partner* (i.e., the technology and the provider) as well as of (2) the *interaction* itself (i.e., perceived interactivity and output quality).

Chapter 2 focuses on the role of perceived task-related characteristics of the *technology as an interaction partner*. In contrast to earlier work assessing task-related characteristics primarily as a unidimensional concept (i.e., competence), this chapter builds on the action regulation theory (Hacker, 1998) to investigate the role of different levels of perceived human-like competencies for disclosure towards AI technologies – using the example of conversational

AI. Two correlational survey studies and three experimental vignette studies highlight the importance of perceived competencies for disclosure: non-users get concerned about privacy and are unwilling to share personal data when conversational AI is perceived as highly competent (i.e., using meta-cognitive heuristics). In contrast, users seem to be willing to share personal information in exchange for intellectual competencies. Although meta-cognitive heuristics were weakly related to more privacy concerns for users, this did not stop them from sharing their personal data. Accordingly, this chapter emphasizes that the perceived characteristics of the technology as interaction partner impact people's disclosure.

Chapter 3 takes a different perspective by focusing on a central aspect of the *interaction* process, namely *perceived interactivity*, for disclosure towards AI technologies. Once more taking the example of conversational AI, two cross-sectional survey studies show that the three different facets of interactivity (cf. Y. Liu, 2003) – namely, active control, reciprocal interaction, and synchronicity – are all positively correlated with disclosure. However, reciprocal interaction seems to be the most relevant interactivity facet. In doing so, this chapter highlights the relevance of perceived interactivity as one central characteristic of the interaction process for understanding disclosure.

Chapter 4 focuses on both perceptions of the *interaction partner* and the *interaction*. Adding to the perspective on the technology itself as interaction partner (addressed in Chapter 3), Chapter 4 considers the role of the technology provider as an omnipresent, but often less salient, interaction partner in human-AI interactions. In an experimental study, where users interacted with an algorithm-based recipe guide, we investigated the interplay of perceived *output quality* (as a characteristic of the interaction) and *provider trustworthiness* (as a characteristic of the interaction partner) on disclosure. The results indicated that people were more willing to share personal information when they perceived the recipe guide's output as being of higher quality and when a more trustworthy company provided the recipe guide. However, there was no interaction between these two factors – indicating that high output quality cannot compensate for low provider trustworthiness and vice versa. Hence, this chapter underlines the distinct importance of perceptions of the interaction partner as well as of the interaction for disclosure.

Chapters 2, 3, and 4, as well as the majority of the existing literature, focus on the perspective of people directly interacting with AI systems to understand disclosure (i.e., people using the system). However, as mentioned at the beginning of this dissertation, there are several situations, in which the individual only has limited ability to decide about data disclosure. For example, suppose a company decides to use AI in the hiring process. In that case, the individual can only decide whether to apply for a job but not whether their personal data will be shared

with an AI system once they submit their application. Thus, it is crucial to consider the perspective of people deciding about AI usage in contexts where it affects large groups of people (e.g., the work context). Notably, decision-makers' perspectives might be essentially different from those of people directly interacting with the AI. For instance, in the work context, decision-makers (i.e., managers) most likely aim more at cost reduction and efficiency and are less concerned about job losses than employees.

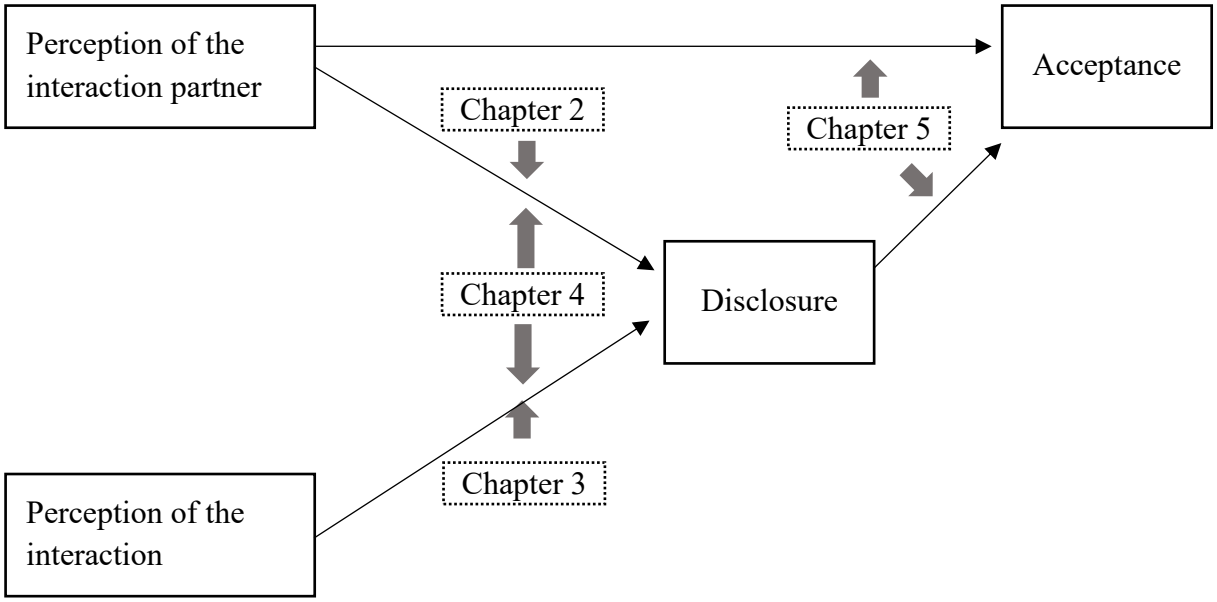
Furthermore, the previous example of AI application in the work context highlights that disclosure in human-AI interactions is not limited to the private use context (as investigated in Chapters 2, 3, and 4). Instead, it is also necessary to better understand information disclosure in other use contexts, such as the work context. Research in this context is of particular relevance as AI is used by more and more companies (Kim-Schmid & Raveendhran, 2022). In addition, AI implementation in the workplace is expected to drastically impact jobs, workforce structure, decision-making, knowledge management, and the organization as a whole (Glikson & Woolley, 2020).

Thus, *Chapter 5* aims at broadening the findings of the other empirical chapters by focusing on the work context and by investigating the perspective of decision-makers (i.e., managers). In order to do so, the research explores decision-makers' perspectives on the relevance of personal data *disclosure* and perceptions of the *interaction partner* (i.e., the AI's perceived *functionality*) for AI acceptance. More precisely, this chapter addresses the relevance of disclosure by comparing a business area dealing particularly with the personal data of employees and applicants (i.e., human resources) to other business areas (i.e., finances, marketing). Besides, the research presented in this chapter aims at an initial understanding of the role of AI functionality (i.e., what AI is capable of doing, as a task-related characteristic of the interaction partner) for managers' acceptance. As competencies in Chapter 3, AI functionality was assessed on different hierarchical levels to get a more fine-grained understanding of its effects. Within three experimental vignette studies, managers were more skeptical about using AI in human resources than in other business areas (i.e., finances and marketing). Furthermore, the studies in this chapter give first insights into the impact of AI functionality for disclosure by indicating that too high functionality (i.e., consequential implementation without a-priori human control) is (currently) not accepted. In doing so, this chapter broadens the perspective of this dissertation by considering not only users of AI technology but also managers as important stakeholders in work-related use contexts. By addressing perceptions of the AI as interaction partner as well as the role of information disclosure for AI acceptance, Chapter 5 extends the empirical insights

of this dissertation. An overview of all relationships investigated in the current dissertation is depicted in Figure 1. [3]

Figure 1

Overview of The Relationships Tested in The Current Dissertation



[3] Please note that the four empirical chapters of this dissertation (i.e., Chapters 2-5) are written in a way that allows for reading them independently. As they are in parts based on similar theoretical and empirical foundations, there might be some overlap between the empirical chapters as well as with the overall introduction of this thesis.

The following chapter contains a manuscript that is the result of a cooperation between Miriam Gieselmann (first author) and Prof. Dr. Kai Sassenberg (second author). The following table depicts the contributions of the PhD candidate (and of the co-author, respectively) to the manuscript as well as the current status in the publication process:

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
Miriam Gieselmann	1	60	80	70	70
Kai Sassenberg	2	40	20	30	30
Title of Paper:		The more competent, the better? – The effects of perceived competencies on disclosure towards conversational Artificial Intelligence			
Status in publication process		Content identical version published as: Gieselmann, M., & Sassenberg, K. (2023). The more competent, the better? – The effects of perceived competencies on disclosure towards conversational Artificial Intelligence. *Social Science Computer Review, 41*(6), 2342-2363. https://dx.doi.org/10.1177/08944393221142787			

Chapter 2: Characteristics of the Interaction Partner

During the last years, artificial intelligence (AI), and especially conversational AI systems, such as Amazon Alexa or Google Assistant, have become ubiquitous. Worldwide, 3.25 billion conversational AI systems (also called voice assistants, personal intelligent agents, etc.) had been in use in 2019 (Statista, 2022). Conversational AI offers several benefits to its users, such as convenience, utility, enjoyment, or personalization and can assist its users in a wide range of tasks (e.g., access to information, entertainment, online shopping, control of smart home devices).

To offer these functionalities and particularly to adapt the interaction experience to the user and their individual needs, conversational AI needs to gather personal information. For instance, location-based requests can only be answered adequately if the system has access to the user's current location, a system can only provide its user with news fitting their interests if it has information about the individual's interests, daily routines can only be created if the system has the necessary information about the user's habits, and the most important information from mails and appointments can only be extracted if the system has access to the mailing and calendar applications of the user. This implies that conversational AI collects a significant

amount of the user's personal data which is often considered problematic: users as well as non-users of conversational AIs report concerns about privacy and the amount of data collected by conversational AI (e.g., Dubiel et al., 2018; Liao et al., 2019).

In this context, it is important to understand which factors guide people's disclosure towards technologies. Up to now, research has investigated associations between disclosure and trust in the technology provider (e.g., Joinson et al., 2010; Pal et al., 2020), individual user characteristics (e.g., Bansal et al., 2010; Mohamed & Ahmad, 2012), as well as objective system characteristics (e.g., Easwara Moorthy & Vu, 2015) such as anthropomorphic design features (e.g., Ha et al., 2021; Lucas et al., 2014) or functionality.

Research on system functionality has often focused on objective system capabilities (e.g., Bandara et al., 2020; Ha et al., 2021). Yet, an approach that has to the best of our knowledge not been taken to understand disclosure, is to consider perceptions of conversational AI's functionality or competence from a more human-oriented perspective. This is surprising because this approach is particularly promising, given that people tend to treat technical systems as social actors (Nass & Moon, 2000; Reeves & Nass, 1996) and anthropomorphize them (i.e., ascribe them human-like characteristics; Epley et al., 2007). Plus, conversational AI is far more capable of communicating in a human-like manner than earlier technologies and it has been observed that some users even engage in small talk with it (Purington et al., 2017). Overall, this suggests that the human-oriented perspective is most likely in line with how people approach conversational AI.

Thus, the present research investigates the associations between perceptions of human-like competence in conversational AI and disclosure towards this technology. In doing so, we aim to contribute to the understanding of people's disclosure towards technologies and shed light on the impact of perceived system competencies.

Technology Competence can Heighten Acceptance and Disclosure

At first sight, more functionality of technical systems seems very positive and should, accordingly, increase the perceived usefulness as well as the adoption and acceptance of the respective technologies. Several theoretical models include this aspect as a core driver of technology adoption (e.g., perceived usefulness in the Technology Acceptance Model, TAM; Davis et al., 1989; performance expectancy in the Unified Theory of Acceptance and Use of Technology, UTAUT; Venkatesh et al., 2003). Furthermore, empirical research has repeatedly shown that functionality of technical systems is highly relevant for technology acceptance, and adoption of technologies in general and conversational AI in specific (e.g., McLean & Osei-Frimpong, 2019; Moriuchi, 2019; Moussawi et al., 2021; Pitardi & Marriott, 2021; Shao &

Kwon, 2021). Besides, for people who already use a (smart) technology, competence is an important driver of usage continuance intentions (e.g., Dehghani, 2018; Moussawi et al., 2022).

More relevant to the current research question, functionality can also have a positive impact on users' disclosure towards a technology. Several studies have shown that this willingness increases as people perceive to get benefits in exchange: perceived benefits (e.g., increased personalization or usefulness) can outweigh privacy concerns and lead to a heightened information disclosure (Chellappa & Sin, 2005; Sharma & Crossler, 2014; Xu et al., 2013) and, in some cases, users accept privacy intrusions for convenience in smart home technologies (Townsend et al., 2011).

Beyond the context of conversational AI, there is research on the relation between a human-like competence measure of technologies (i.e., assessing perceived competence based on the Stereotype Content Model by Fiske et al., 2002) and acceptance. In several studies, higher competence perceptions were associated with a higher acceptance of technologies: for instance, a higher believability of these agents (Demeure et al., 2011) and a higher willingness to use (X. S. Liu et al., 2022). There is also very first evidence regarding conversational AI: Pitardi and Marriott (2021) report that, besides perceived usefulness, also higher subjective perceptions of competence were related to a more positive attitude towards and higher trust in conversational AI.

In summary, the research described so far seems to suggest that functionality and competence of technical systems is key to adoption and acceptance. This research has either focused on the technical functionality of systems or, when considering a human-like consideration of competence, used a very simplified approach with a single dimension.

But: Technology Competence can Also Lessen Acceptance and Disclosure

However, in contrast to the previously described findings, there are also indications that increased functionalities might not always be desirable. Technological progress and increased system capabilities are often assumed to be associated with a general fear of privacy invasion and exploitation of personal data, especially in the field of AI (e.g., Bandara et al., 2020; Ha et al., 2021). Further, referring to an 'uncanny valley of mind', it was observed that AI described as having a human-like mind in comparison to basic algorithms as a system basis led to participants feeling more eeriness (Stein et al., 2020).

Empirical studies focusing on human-like competencies did not only find positive effects as the ones described above but also detrimental effects of technology competence. Złotowski et al. (2017) reported that autonomous robots were perceived as more threatening and evoked stronger negative attitudes than non-autonomous ones - thereby suggesting that

competencies are not always seen as merely positive. Similarly, McKee et al. (2022) found that perceptions of higher competence predicted lower preferences for cooperative agents – with the effect being stronger than and opposed to the effect of objective performance measures, thus, highlighting the important role of perceived competencies.

The previously described findings are focused on detrimental effects regarding perceptions of eeriness and general usage decisions. Few studies considered the (detrimental) effects of competence on disclosure. As an exception, Peng et al. (2019) conducted a study in which participants interacted with a robot recommending them items. They report that participants shared more feelings and thoughts about the recommended items with a robot who had a medium proactivity compared to a robot with either low or high proactivity. Here, obviously one specific aspect of competence resulted in detrimental effects. In another study, Liu et al. (2021) observed that higher perceptions of competence in a social robot led to more privacy concerns.

In sum, the findings presented in this and the preceding section are inconsistent. Competence can be positive or detrimental to technology acceptance. We sought to integrate these findings by taking a more differentiated approach to competence. We argue that it will, for example, potentially make a large difference whether users perceive a technical system as capable of completing tasks by following given rules or as being able to flexibly pursue their own independent goals.

Differentiation of Competence

Four levels of competencies of conversational AI can be distinguished from a human-oriented perspective by applying the theory of human action regulation (Frese & Zapf, 1994; Hacker & Sachse, 2014; Semmer & Frese, 1985). The action regulation theory – originally introduced by Hacker (1998) to describe different levels of action regulation in human work actions – initially included three levels and was extended to four levels by Semmer and Frese (1985). These levels differ in the amount of required action regulation.

On the lowest *sensorimotor level*, highly automatized movement patterns and cognitive routines are regulated (Zacher, 2017). It is assumed that actions on this level are not associated with independent and conscious goals but are usually triggered by regulation processes at higher levels (Zacher, 2017). Applied to the perception of a technology's competence, this level includes acting according to existing (or pre-defined) rules to solve clearly defined problems and executing specific commands.

The second *level of flexible action patterns* includes action regulation based on (semi-)conscious automatized schemata or scripts: humans process information from the

environment according to well-established rules. Signals from the environment can then activate action patterns (Zacher, 2017). For technologies, flexible action patterns can be characterized as an adaptability to different factors (e.g., situation, environment, or task). We expected that these lower two levels are commonly perceived in technologies. Thus, we were unsure whether their perception would have a meaningful impact on people's disclosure towards conversational AI.

Regulating actions on the *intellectual level* implies the regulation of new or complex actions, entailing the development and selection of goals and detailed action plans (Zacher, 2017). Action regulation on this level requires attention and cognitive effort as it includes the analysis and evaluation of novel and complex information (Zacher, 2017). For technologies, intellectual competencies include anticipating and planning, finding innovative solutions as well as dealing with complex or incomplete information. We expected that people would perceive these intellectual competencies as clearly beneficial and useful. As the system uses the submitted data directly and it seems to be clear what the data is used for, we expected that higher intellectual competencies would heighten disclosure towards conversational AI.

Finally, the *level of meta-cognitive heuristics* involves the use of more abstract, less task-oriented templates, strategies, and heuristics to guide action regulation enabling to solve similar problems in an efficient and effective way (Frese & Zapf, 1994; Semmer & Frese, 1985; Zacher, 2017). For technologies, meta-cognitive heuristics include learning as well as the development and adaptation of universal strategies based on previous events and interactions. We expected that people would perceive meta-cognitive heuristics as autonomous and strategic – attributes that might be perceived as too competent and, thus, eerie (Stein et al., 2020; Złotowski et al., 2017). Accordingly, we expected that meta-cognitive heuristics would lessen disclosure towards conversational AI.

Differentiation of Disclosure

Regarding the assessment of disclosure, we distinguished between (a) willingness to disclose as an attitude regarding behavioral intention and (b) privacy concerns as an indicator for more abstract attitudes towards privacy. We chose this distinction because previous research has indicated that high privacy concerns do not necessarily stop people from sharing their data ('privacy paradox', for a review see Kokolakis, 2017). To investigate potential different effects of competencies on these two aspects of disclosure, we differentiated them in our research question and hypotheses accordingly:

RQ: Do lower order competencies (sensorimotor competencies, flexible action patterns) of a conversational AI affect (a) willingness to disclose and (b) privacy concerns?

H1: Higher intellectual competencies of a conversational AI predict (a) a higher willingness to disclose and (b) lower privacy concerns.

H2: Higher meta-cognitive heuristics of a conversational AI predict (a) a lower willingness to disclose and (b) higher privacy concerns.

Current Research

We tested these ideas in five online studies. In the first study, we aimed to develop a scale for the assessment of competencies according to the action regulation theory across different technologies. Further, we exploratory investigated associations between disclosure and the four levels of competence in conversational AI. As the action regulation theory for human working actions had, up to now, not been used to assess the perception of competencies in technological systems, we did not initially formulate the described hypotheses but developed them based on the described theoretical considerations in combination with the results of this first study. Starting with Study 1.2 the hypotheses were preregistered. All deviations from the preregistrations are reported in detail in Appendix A.

In the second study, we aimed to validate and improve the scale as well as to replicate the findings regarding intellectual competencies and meta-cognitive heuristics in conversational AI. In Studies 1.3 to 1.5, we experimentally manipulated intellectual competencies and meta-cognitive heuristics to check for the causality of the observed effects.

In the research process, it became obvious that effects might differ based on usage experience, which is in retrospect highly plausible. As users have, in contrast to non-users, actual interaction experience with conversational AI, its competencies could have a different meaning for them. Accordingly, users might focus more on the positive aspects rather than the negative ones. Thus, we clearly indicate the usage experience of our samples in all studies and discuss its potential effects in the discussions of the respective studies and the overall research.

Study 1.1

Method

Design and Participants. We aimed to sample 300 valid cases for this study (https://aspredicted.org/tq3m3.pdf) in order to reach stable correlations (Schönbrodt & Perugini, 2013). We recruited participants via Prolific for a 10-min online survey remunerated with £1.25. Participants were randomly assigned to one of five conditions. Depending on the condition, participants were surveyed about their perception of one specific technology (i.e., either conversational AI, streaming service, search engine, autonomous vehicle, or washing machine).

After exclusion as pre-registered (for details, see Appendix A), the eligible sample consisted of N = 358 participants (59.8% male, 38.5% female, 1.4% non-binary; aged: M = 25.5, 18-62 years). In the conversational AI condition, the final sample consisted of 72 participants (54.2% male, 44.4% female, 1.4% non-binary; aged: M = 27.0, 18-56 years; 27 using conversational AI maximum once a month, 21 at least several times a week and 24 with usage frequencies in between).

Procedure. Participants were randomly assigned to one of the five conditions. In each condition, participants first read a short description of the respective technology: either a conversational AI or another technology (i.e., washing machine, search engine, streaming service, autonomous vehicle). This was done to increase the variance of the responses to the competence scales. Therefore, we used all conditions for the scale construction, but only the conversational AI condition to test the relation between the competencies and disclosure. Then, participants were surveyed about their perceptions of the respective technology's competencies, their privacy concerns, and willingness to disclose. A complete list of all measures taken, instructions, and items is provided in Appendix A – this applies for all studies reported in this paper.

Measures.

Action Regulation Competencies. To assess competencies on the levels of the action regulation theory, we developed 24 items based on the level descriptions by Zacher (2017). All scales were assessed on seven-point scales (1 = *strongly disagree*, 7 = *strongly agree*). Based on the theory, we assumed a four-factor structure and conducted an initial exploratory factor analysis (see Appendix, Table A1). We excluded items with ambiguous first loadings, substantial cross loadings and/or loading highest on a factor they were not meant to capture, except for two items which were initially intended for meta-cognitive heuristics but included in the intellectual competencies scale based on the EFA results and reconsideration of their content. Afterwards, we conducted a second EFA with the remaining items. The results confirmed the expected factor structure (see Appendix, Table A6). Example items for the resulting scales are: '… acts according to predefined rules.' (sensorimotor competencies), 'In different situations, … behaves differently.' (flexible action patterns), '… can plan actions' (intellectual competencies), or '… adapts its behavior based on prior events' (meta-cognitive heuristics). Descriptive results for the overall sample as well as the conversational AI subsample for all scales are reported in Table 1.

Table 1

Scale Reliabilities, Means, and Standard Deviations for Competencies & Disclosure for Full Sample of Study 1.1 (N = 358), Conversational AI Subsample of Study 1.1 (n = 72), and Study 1.2 (N = 334)

	Study 1.1							Study 1.2			
	Full sample (N = 358)				Conversational AI (n = 72)						
Scale	k	α	M	SD	α	M	SD	k	α	M	SD
Competence level											
Sensorimotor	5	0.71	5.54	0.78	0.68	5.56	0.73	6	0.70	5.59	0.70
Flexible action patterns	3	0.62	4.93	1.08	0.63	4.60	1.02	6	0.79	4.20	1.06
Intellectual	4	0.71	3.90	1.18	0.66	3.76	1.04	6	0.70	3.89	0.93
Meta-cognitive heuristics	3	0.77	4.41	1.42	0.40	4.95	0.87	6	0.76	4.40	0.94
Disclosure											
Privacy concerns	4	0.86	4.06	1.40	0.80	4.49	1.25	5	0.90	4.03	1.35
Willingness to disclose	5	0.88	3.48	1.46	0.90	3.60	1.46	5	0.90	3.57	1.46

Note. k = number of items included in scale

Disclosure. *Privacy concerns* were assessed with four items (adapted from a scale called perceived legitimacy by Alge et al., 2006, e.g. 'I feel that ...'s practices are an invasion of privacy.'). *Willingness to disclose* was measured with three items (adapted from Gerlach et al., 2015, e.g., 'I would provide a lot of information to ... about things that represent me personally.') and two additional, self-developed items.

Results

To explore the relation between perceived competencies of conversational AI and disclosure, we separately regressed (1) willingness to disclose and (2) privacy concerns towards conversational AI to the four levels of competence. Willingness to disclose towards conversational AI was neither predicted by the sensorimotor level (β = 0.13, p = .222, 95%-CI[-0.08, 0.34]) nor by flexible action patterns (β = 0.12, p = .315, 95%-CI[-0.11, 0.35]). However, higher intellectual competencies predicted a higher willingness to disclose (β = 0.45, p < .001, 95%-CI[0.23, 0.66]), whereas higher meta-cognitive heuristics predicted a lower willingness to disclose (β = -0.24, p = .036, 95%-CI[-0.46, -0.02]).

Privacy concerns regarding conversational AI were neither predicted by the sensorimotor level (β = -0.16, p = .148, 95%-CI[-0.39, 0.06]) nor by flexible action patterns

(β = 0.04, p = .733, 95%-CI[-0.20, 0.29]). However, higher intellectual competencies predicted less privacy concerns (β = -0.36, p = .003, 95%-CI[-0.59, -0.13]), whereas higher meta-cognitive heuristics predicted more privacy concerns (β = 0.25, p = .039, 95%-CI[0.01, 0.48]).

Discussion

In Study 1.1, neither willingness to disclose nor privacy concerns towards conversational AI were associated with perceived competencies on the two lower levels (i.e., sensorimotor competencies and flexible action patterns). Higher intellectual competencies were associated with more willingness to disclose and less privacy concerns, whereas higher meta-cognitive heuristics were associated with less willingness to disclose and more privacy concerns. These results are in line with Hypotheses 1 and 2, which we pre-registered for Studies 1.2-1.5.

The exclusion of several items of the action regulation scales resulted in a low number of items for some scales. Besides, the internal consistency for some of the scales were non-satisfactory. Moreover, the sample focusing on perceptions of conversational AI was rather small and consisted in large parts of conversational AI non-users. Therefore, the results of Study 1.1 should be interpreted with caution. To overcome these limitations, we developed new items to assess competencies on the four levels of action regulation aiming to improve the reliability of the scales and recruited a larger sample size of actual conversational AI users to assess their perception of this technology in Study 1.2.

Study 1.2

Method

Design, Participants, and Procedure. We conducted a correlational study (https://aspredicted.org/v87ub.pdf). We aimed to collect 340 valid cases to reach a power of 95% (alpha = .05) for observing effects with a minimum effect size of f^2 = 0.03 in a multiple regression analysis with four predictors. For this purpose, we recruited participants owning a home assistant or smart hub via Prolific for a 12-min online study remunerated with £1.50. 425 participants who matched the pre-screening (i.e., owning home assistant / smart hub and having used a conversational AI before) completed the survey. After exclusion as pre-registered, the eligible sample consisted of N = 334 participants (54.2% male, 44.0% female, 1.8% non-binary; age: M = 30.3, 18-73 years).

Study 1.2 was a conceptual replication of Study 1.1 with the following changes: (1) we focused exclusively on the perception of conversational AI and (2) included only conversational AI users as participants.

Measures.

Action regulation competencies. Action regulation competencies were assessed with the items included in the final scales of Study 1.1 and newly developed items resulting in six items per competence level. We conducted confirmatory factor analyses for both a three-factor structure and the initially assumed four-factor structure using the R package "lavaan" (Rosseel et al., 2021) (see Table 2, all latent factors were correlated, no further error correlations assumed). Model fit statistics indicated an acceptable fit for both models. However, the fit was slightly better for the four-factor solution, $CFI = 0.86$, $RMSEA = 0.060$, 90%-CI_{RMSEA}[0.053, 0.067], $SRMR = 0.077$, compared to the three-factor solution, $CFI = 0.83$, $RMSEA = 0.065$, 90%-CI_{RMSEA}[0.059, 0.072], $SRMR = 0.078$, $\Delta\chi^2(3) = 59$. Thus, we stuck to the originally planned four scales of action regulation competencies. Descriptive results and internal consistencies for all scales are reported in Table 1.

Table 2

Results of a Confirmatory Factor Analysis of the Action Regulation Items (Study 1.2: N = 334)

Latent variable	Factor loading	
Item	*Est.*	*SE*
Sensorimotor competencies		
... solves clearly defined problems in a specific domain.	1.00	
... acts according to predefined rules.	0.64	0.09
... considers information that was explicitly submitted for task completion.	0.68	0.10
... behaves in a preprogrammed manner.	0.41	0.08
... executes specific commands.	0.62	0.09
... is prepared to deal with specific situations.	0.84	0.12
Flexible action patterns		
... adjusts its behavior depending on situational factors.	1.00	
In different situations, ... behaves differently.	0.90	0.08
...'s behavior differs depending on the given task.	0.79	0.08
...'s behavior depends on the given context.	0.75	0.08
... adapts its behavior depending on who is interacting with it.	0.97	0.09
...'s behavior is sensitive to environmental factors.	0.71	0.09
Intellectual competencies		
... anticipates potential problems.	1.00	
... can deal with incomplete information.	0.94	0.10

Latent variable	Factor loading	
Item	*Est.*	*SE*
... solves unknown tasks.	0.71	0.10
... finds innovative solutions.	0.80	0.10
... can plan actions.	0.77	0.11
... can extract key information from complex input.	0.67	0.09
Meta-cognitive heuristics		
... learns from mistakes.	1.00	
... adapts its behavior based on prior events.	0.88	0.07
... transfers knowledge to other domains.	0.33	0.06
Based on previous incidents, ... derives universal strategies.	0.50	0.06
... develops general strategies to accomplish various tasks.	0.53	0.07
... uses information from prior occurrences to solve related tasks.	0.91	0.06
Covariances	*r*	*p*
Sensorimotor competencies		
Flexible action patterns	0.16	.034
Intellectual competencies	0.10	.176
Meta-cognitive heuristics	0.28	< .001
Flexible action patterns		
Intellectual competencies	0.63	< .001
Meta-cognitive heuristics	0.64	< .001
Intellectual competencies		
Meta-cognitive heuristics	0.76	< .001

Note. '…' was replaced with 'the voice assistant'.

Disclosure. Expect for the change of one item in the *privacy concerns* scale ('I am concerned about my privacy when using the voice assistant.') scales were the same as in Study 1.1.

Results

We tested the predictions that higher intellectual competencies would predict a higher willingness to disclose (H1a), whereas higher meta-cognitive heuristics would predict a lower willingness to disclose (H2a). Supporting H1a, a multiple regression analysis (pre-registered) showed that higher intellectual competencies were associated with more willingness to disclose ($\beta = 0.22$, $p = .001$, 95%-CI[0.09, 0.35]). H2a was not supported, as meta-cognitive heuristics

did not predict willingness to disclose (β = -0.05, p = .430, 95%-CI[-0.19, 0.08]). Further, higher willingness to disclose was predicted by higher sensorimotor competencies (β = 0.22, p < .001, 95%-CI[0.11, 0.32]) but not by flexible action patterns (β = 0.09, p = .155, 95%-CI[-0.03, 0.21]).

Regarding privacy concerns, we predicted that higher intellectual competencies would predict lower privacy concerns (H1b), whereas higher meta-cognitive heuristics would predict higher privacy concerns (H2b). A multiple regression analysis (pre-registered) supported H1b as higher intellectual competencies (β = -0.14, p = .040, 95%-CI[-0.28, -0.01]) predicted lower privacy concerns and also H2b as higher meta-cognitive heuristics (β = 0.15, p = .034, 95%-CI[0.01, 0.29]) predicted higher privacy concerns. Furthermore, the analysis showed that lower privacy concerns were predicted by higher sensorimotor competencies (β = -0.19, p = .001, 95%-CI[-0.30, -0.08]) but not by flexible action patterns (β = 0.05, p = .400, 95%-CI[-0.07, 0.18]).

Discussion

In both studies, higher perceived intellectual competencies were associated with higher willingness to disclose and less privacy concerns, supporting H1a and H1b. Regarding the detrimental relation between meta-cognitive heuristics and disclosure, the results were somewhat inconsistent. Support for its negative relation with willingness to disclose (H2a) was only found in Study 1.1, but not in Study 1.2. However, higher perceived meta-cognitive heuristics were in both studies associated with more privacy concerns as suggested in H2b.

In a next step, we aimed to test whether the differences between the results of both studies result from the fact that Study 1.2 relied on users, whereas participants of Study 1.1 had mixed usage experience. Thus, we additionally included usage experience as a quasi-experimental factor in Study 1.3. Moreover, both studies offer only offer correlative support for the hypotheses – thus not allowing conclusions about causality. Studies 1.3-1.5, therefore, experimentally manipulated the conversational AI's intellectual competencies and meta-cognitive heuristics to check for causality of the observed relations and to make sure that the effects are due to one or the other concept (i.e., both concepts are independent). Furthermore, Study 1.1 and 1.2 did not show a clear pattern of associations between sensorimotor competencies and flexible action patterns and disclosure. Therefore, we did not further focus on these two competence levels (and respectively the RQ) in the following studies.

Study 1.3

Method

Design and Participants. Study 1.3 was implemented as a 2 (intellectual competencies: low vs high) x 2 (meta-cognitive heuristics: low vs high) x 2 (usage experience: users vs non-users) between-subjects design. Usage experience was a quasi-experimental factor: participants were regarded as users if they indicated to use a voice assistant at least several times a week and as non-users if they indicated to use it maximum once a month. Participants were randomly assigned to one of the four conditions made up by the other two factors.

We aimed to collect 344 valid cases in order to reach a power of 80% ($\alpha = 0.05$) for observing a minimum effect of $f^2 = 0.02$ in a MANOVA with eight groups and two dependent variables. We recruited participants via Prolific for an 11 min-online study remunerated with £1.40. After exclusion as pre-registered (https://aspredicted.org/qg5m4.pdf), the eligible sample consisted of $N = 323$ participants (35.3% male, 63.8% female, 0.9% non-binary; aged: $M = 39.3$, 18-81 years), with $n = 167$ non-users and $n = 156$ users.

Procedure. Participants read a short description of a fictional planned update for an existing conversational AI. Within this description, sensorimotor competencies and flexible action patterns were consistently described as improved compared to before the update as follows: *'After the update, the voice assistant will have a substantially improved capability to react to more different user requests by executing specific commands. Another improvement is that it will consider characteristics of specific situations more than before (e.g., behave differently depending on where and when the interaction takes place, or who is interacting with it).'*

The description of intellectual competencies and meta-cognitive heuristics differed depending on the experimental condition. In the *high intellect* conditions, intellectual competencies were described as improved after the update by the following description and an additional example (see Appendix A) to illustrate the respective competencies: *'Besides, after installing the update, the voice assistant will have improved its capabilities to gather key information out of complex input, and to plan actions in advance, and to find innovative solutions'.* In contrast, in the *low intellect* conditions, intellectual competencies were described as not improved by the update using the respective description and the same example to illustrate the lacking competencies: '*Yet, after installing the update, the voice assistant will NOT have improved its capabilities to gather key information out of complex input, nor to plan actions in advance, nor to find innovative solutions.'*

Similarly, meta-cognitive heuristics were described as improved by the update in the *high meta* conditions with the following description and an illustrating example: *'However, the*

voice assistant will have improved its capabilities to use information from prior events to solve related tasks, and to develop general strategies for task completion, and to learn from its previous mistakes.' In the *low meta* conditions, a similar text was used to indicate that the respective competencies were not improved by the update.

After reading the update description, participants were asked about their disclosure to the conversational AI after the described update followed by a manipulation check, and a measure of conversational AI usage.

Measures. *Privacy concerns* (α = 0.94) and *willingness to disclose* (α = 0.91) were assessed with the same scales as in Study 1.2. *Manipulation checks* consisted of three items each for intellectual competencies (α = 0.78) and meta-cognitive heuristics (α = 0.70). Close care was taken that the wording of the manipulation and the manipulation checks did not overlap.

Results

Manipulation Check. As intended, participants in the high intellect conditions ($M = 4.23$, $SD = 1.19$) perceived significantly higher intellectual competencies compared to participants in the low intellect conditions ($M = 2.72$, $SD = 1.07$), $F(1, 315) = 181.47$, $p < .001$, $\eta_p^2 = 0.37$. Further, participants in the high intellect conditions ($M = 4.21$, $SD = 1.48$) also perceived significantly higher meta-cognitive heuristics compared to participants in the low intellect conditions ($M = 3.79$, $SD = 1.42$), $F(1, 315) = 13.26$, $p < .001$, $\eta_p^2 = 0.04$.

Participants in the high meta conditions perceived intellectual competencies ($M = 3.98$, $SD = 1.34$) as higher than participants in the low meta conditions ($M = 3.01$, $SD = 1.22$), $F(1,315) = 74.68$, $p < .001$, $\eta_p^2 = 0.19$, and also, as intended, meta-cognitive heuristics ($M = 5.05$, $SD = 0.74$) as higher compared to those in the low meta conditions ($M = 2.93$, $SD = 1.23$), $F(1, 315) = 371.48$ $p < .001$, $\eta_p^2 = 0.54$.

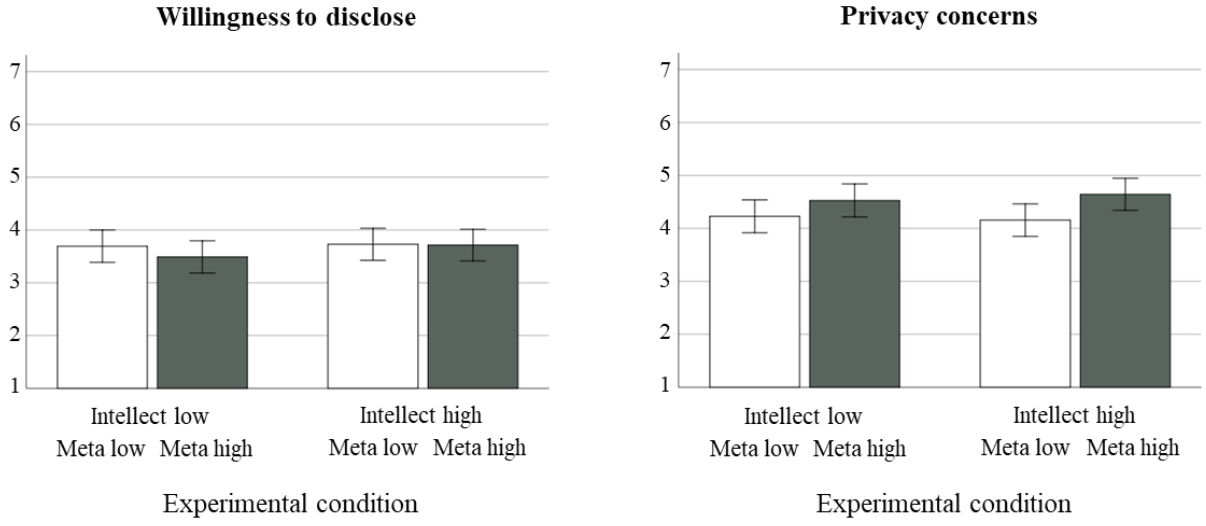
There was no main effect of usage, $p = .398$. However, a significant meta by usage interaction indicates that the spill-over effect of the meta manipulation on perceived intellectual competencies was higher for users compared to non-users, $F(1, 315) = 6.73$, $p = .010$, $\eta_p^2 = 0.02$, while no interaction occurred for perceived meta-cognitive heuristics, $p = .624$. Further, the interaction between the two experimental factors indicates that this spillover effect on intellectual competencies was stronger in the high intellect compared to the low intellect conditions, $F(1, 315) = 4.71$, $p = .031$, $\eta_p^2 = 0.01$, but absent for perceived meta-cognitive heuristics, $p = .611$. Accordingly, both manipulations had the intended effects, but they also had considerably smaller effects on the respective other competence level. Given that effects were

substantially larger on the intended level of competence and also the observed interactions were small in effect size, we considered the manipulation successful.

Hypotheses Testing. A MANOVA (pre-registered) with privacy concerns and willingness to disclose as dependent variables revealed significant effects of usage, $F(2, 314) = 12.12$, $p < .001$, $\eta_p^2 = 0.07$, Wilk's $\lambda = 0.93$, and the meta manipulation, $F(2, 314) = 4.34$, $p = .014$, $\eta_p^2 = 0.03$, Wilk's $\lambda = 0.97$, but not of the intellect manipulation nor any of the interactions between these three factors, all ps > .100, see Figure 2. Non-users ($M = 4.77$, $SD = 1.45$) reported higher privacy concerns than users ($M = 4.01$, $SD = 1.39$), $F(1, 315) = 22.68$, $p < .001$, $\eta_p^2 = 0.07$. Besides, non-users ($M = 3.32$, $SD = 1.49$) reported lower willingness to disclose than users ($M = 3.99$, $SD = 1.27$), $F(1, 315) = 18.39$, $p < .001$, $\eta_p^2 = 0.06$. Further, participants in the high meta conditions ($M = 4.62$, $SD = 1.49$) had higher privacy concerns compared to those in the low meta conditions ($M = 4.19$, $SD = 1.42$), $F(1, 315) = 6.33$, $p = .012$, $\eta_p^2 = 0.02$, whereas the meta manipulation did not have an impact on willingness to disclose, $F(1, 315) = 0.50$, $p = .479$, $\eta_p^2 = 0.00$.

Figure 2

Privacy Concerns and Willingness to Disclose by Experimental Condition (Study 1.3, N = 323)



Note. 95% error bars.

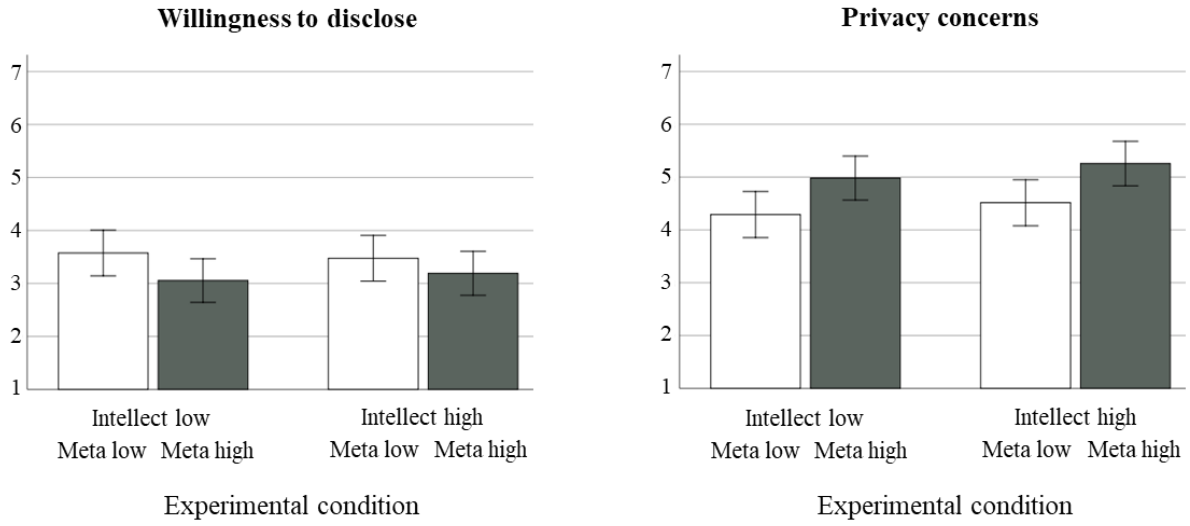
Although there was no significant interaction of any experimental factor with usage, we exploratory analyzed the subsamples of non-users and users separately due to the differences between the results of Study 1.1 and 1.2 by computing MANOVAs with privacy concerns and willingness to disclose as dependent variables.

For non-users, the meta manipulation had a multivariate main effect, $F(2, 162) = 5.80$, $p = .004$, $\eta_p^2 = 0.07$, Wilk's $\lambda = 0.93$, but we found neither an intellect main effect, $p = .218$,

nor an interaction, $p = .683$, see Figure 3. Non-users in the *high meta* condition reported higher privacy concerns ($M = 5.12$, $SD = 1.36$) compared to those in the *low meta* conditions ($M = 4.40$, $SD = 1.45$), $F(1, 163) = 10.76$, $p = .001$, $\eta_p^2 = 0.06$. The willingness to disclose was not affected, $p = .083$. For users, none of the effects turned out significant, all *p*s > .250.

Figure 3

Privacy Concerns and Willingness to Disclose by Experimental Condition in the Non-User Subsample (Study 1.3, n = 167)



Note. 95% error bars.

Discussion

In this experiment, we observed no effect of the intellect manipulation on neither willingness to disclose nor privacy concerns – thus not supporting H1a and H1b. The manipulation of meta-cognitive heuristics had no effect on willingness to disclose, however, for the overall sample and the non-user subsample, there was a marginal trend that participants in the high meta condition were less willing to disclose. However, the manipulation of meta-cognitive heuristics significantly affected participants' privacy concerns. The effect observed in the overall sample was driven by the non-users whereas the effect was non-significant in the user subsample. Accordingly, the results support H2b for non-users but not for users. In order to (1) validate the effect of meta-cognitive heuristics on privacy concerns for non-users, and (2) test again for a potential effect of intellectual competencies, we conducted Study 1.4.

Study 1.4

Method

Design and Participants. The study was implemented as a 2 (intellectual competencies: low vs. high) x 2 (meta-cognitive heuristics: low vs. high) between-subjects design. We aimed to collect 351 valid cases in order to reach a power of 80% ($\alpha = 0.05$) for observing a minimum effect of $f = 0.15$ in an ANOVA with four groups. We recruited non-users of conversational AI via the Clickworker platform for an 11-min online study renumerated with 1.60€. After the pre-registered exclusion (https://aspredicted.org/39en3.pdf), the eligible sample consisted of $N = 253$ participants (54.1% male, 53.4% female, 1.2% non-binary, one participant preferring not to indicate their gender; aged: $M = 39.8$, 18-72 years).

Procedure. The procedure was similar to Study 1.3 with the following change: the experimental manipulation was implemented via a description of a newly developed conversational AI (versus an update for an existing one in Study 1.3) as described in the following.

Participants read a short scenario in which it was described that a friend of them recently bought a new conversational AI and would now tell them about their usage experience. Similar to Study 1.3, competencies on the two lover levels (i.e., sensorimotor level and flexible action patterns) were always described as high. Depending on the experimental condition, intellectual competencies were described as either low or high (e.g., in the *low intellect* condition by the text *'However, the voice assistant is incapable of dealing with incomplete information and it cannot anticipate potential problems or find innovative solutions.'* and an illustrating example) and similarly meta-cognitive heuristics were described as either high or low (e.g., in the low meta condition by the text *'In addition, I realized that the voice assistant does not learn from mistakes and cannot use information from prior events to solve related tasks. The voice assistant is, thus, incapable to develop abstract strategies based on previous incidents.'* and an illustrating example).

Measures. *Privacy concerns* ($\alpha = 0.92$) and *willingness to disclose* ($\alpha = 0.92$) were assessed with the same scales as in the previous studies. *Manipulation checks* for competencies on the intellectual level ($\alpha = 0.70$) and competencies on the level of meta-cognitive heuristics ($\alpha = 0.74$) again consisted of three items each, which were not part of the description of the conversational AI used for the experimental manipulation in this study.

Results

Manipulation Check. As intended, participants in the *high intellect* conditions ($M = 4.52$, $SD = 1.11$) perceived significantly higher intellectual competencies compared to

the *low intellect* condition (M = 3.73, SD = 1.22), $F(1, 249) = 39.45$, $p < .001$, $\eta_p^2 = 0.14$, but did not differ in their perception of meta-cognitive heuristics, $p = .133$. Participants in the *high meta* conditions perceived intellectual competencies (M = 4.72, SD = 1.03) as higher than participants in the *low meta* conditions (M = 3.56, SD = 1.14), $F(1, 249) = 82.56$, $p < .001$, $\eta_p^2 = 0.25$, and also meta-cognitive heuristics (M = 5.55, SD = 0.72) as higher compared to the *low meta* conditions (M = 3.59, SD = 1.25), $F(1, 249) = 234.20$, $p < .001$, $\eta_p^2 = 0.49$. Accordingly, the manipulation of meta-cognitive heuristics spilled over to the perception of intellectual competencies. Nonetheless, effect sizes show that its effect on the perception of meta-cognitive heuristics was substantially larger.

Hypotheses Testing. A MANOVA with privacy concerns and willingness to disclose as dependent variables indicated no significant multivariate effect of the intellect manipulation, $p = .681$, but of the meta manipulation, $F(2, 248) = 3.54$, $p = .030$, $\eta_p^2 = 0.03$, Wilk's $\lambda = 0.97$. The interaction of the two experimental factors was non-significant, $p = .256$. Participants in the *high meta* conditions were less willing to disclose information to the conversational AI (M = 3.39, SD = 1.49) than participants in the *low meta* conditions (M = 3.86, SD = 1.45), $F(1, 249) = 6.14$, $p = .014$, $\eta_p^2 = 0.02$. Further, participants in the *high meta* conditions (M = 4.89, SD = 1.52) had more privacy concerns than participants in the *low meta* conditions (M = 4.43, SD = 1.30), $F(1, 249) = 6.29$, $p = .013$, $\eta_p^2 = 0.03$.

Discussion

The manipulation inducing higher meta-cognitive heuristics led to less willingness to disclose (different from Study 1.3, where we observed no significant effect in the overall sample but a marginal effect in the non-user sample) and more privacy concerns (as in Study 1.3), supporting H2a and H2b.

Similar as in Study 1.3, Study 1.4 again showed no support for an effect of intellectual competencies on willingness to disclose (H1a) or privacy concerns (H1b). To assure that this lack of effect is not due to the lower salience of the intellect manipulation compared to the meta-cognitive heuristics (due to the earlier manipulation in the study process), we conducted Study 1.5 in which we excluded the meta manipulation and, thus, made the intellect manipulation more salient.

Study 1.5

Method

Design and Participants. In large part, Study 1.5 was a conceptual replication of Study 1.4. However, it was implemented as a one factorial (intellectual competencies: low vs high) between-subjects design and meta-cognitive heuristics were not manipulated.

We aimed to collect 486 valid cases in order to reach a power of 80% ($\alpha = 0.05$) in a MANOVA with two groups, one independent and two dependent variables. Participants were recruited via Prolific for an 8 min-online study remunerated with £1.00 following the same pre-screening procedure as in Study 1.3. After exclusion as pre-registered (https://aspredicted.org/947b9.pdf) the eligible sample consisted of $N = 471$ participants (31.4% male, 68.2% female, 0.4% non-binary; aged: $M = 38.7$, 18-79 years).

Procedure and Measures. The procedure was identical to Study 1.4 with the following changes: (1) for the experimental manipulation, we sticked to the description of a newly developed conversational AI but used an illustrating example for the described intellectual competencies similar to the one used in Study 1.3, and (2) we did not manipulate meta-cognitive heuristics.

Measures were identical to Study 1.3: *privacy concerns* ($\alpha = 0.93$), *willingness to disclose* ($\alpha = 0.90$) and the manipulation check for *intellectual competencies* ($\alpha = 0.84$) had good internal consistency.

Results

As intended, perceived intellectual competencies were lower in the low intellect condition ($M = 2.82$, $SD = 1.10$) compared to the high intellect condition ($M = 4.82$, $SD = 0.91$), $t(446.08) = -21.37$, $p < .001$.

A MANOVA with privacy concerns and willingness to disclose as dependent variables indicated no multivariate effect for the intellect manipulation, $F(2, 468) = 1.58$, $p = .207$, $\eta_p^2 = 0.01$, Wilk's $\lambda = 0.99$. Accordingly, there were no differences regarding willingness to disclose information between the *high intellect* ($M = 3.32$, $SD = 1.55$) and the *low intellect* condition *(*$M = 3.20$, $SD = 1.49$), $F(1,469) = 0.79$, $p = .374$, $\eta_p^2 = 0.00$. Privacy concerns did also not differ between the *high intellect* ($M = 5.04$, $SD = 1.46$) and the *low intellect* condition *(*$M = 4.98$, $SD = 1.42$), $F(1, 469) = 0.22$, $p = .643$, $\eta_p^2 = 0.00$.

Discussion

Similar to Studies 1.3 and 1.4, Study 1.5 again showed no effect of the intellect manipulation on neither willingness to disclose information nor privacy concern, thus not supporting H1a and H1b.

Results Across Studies

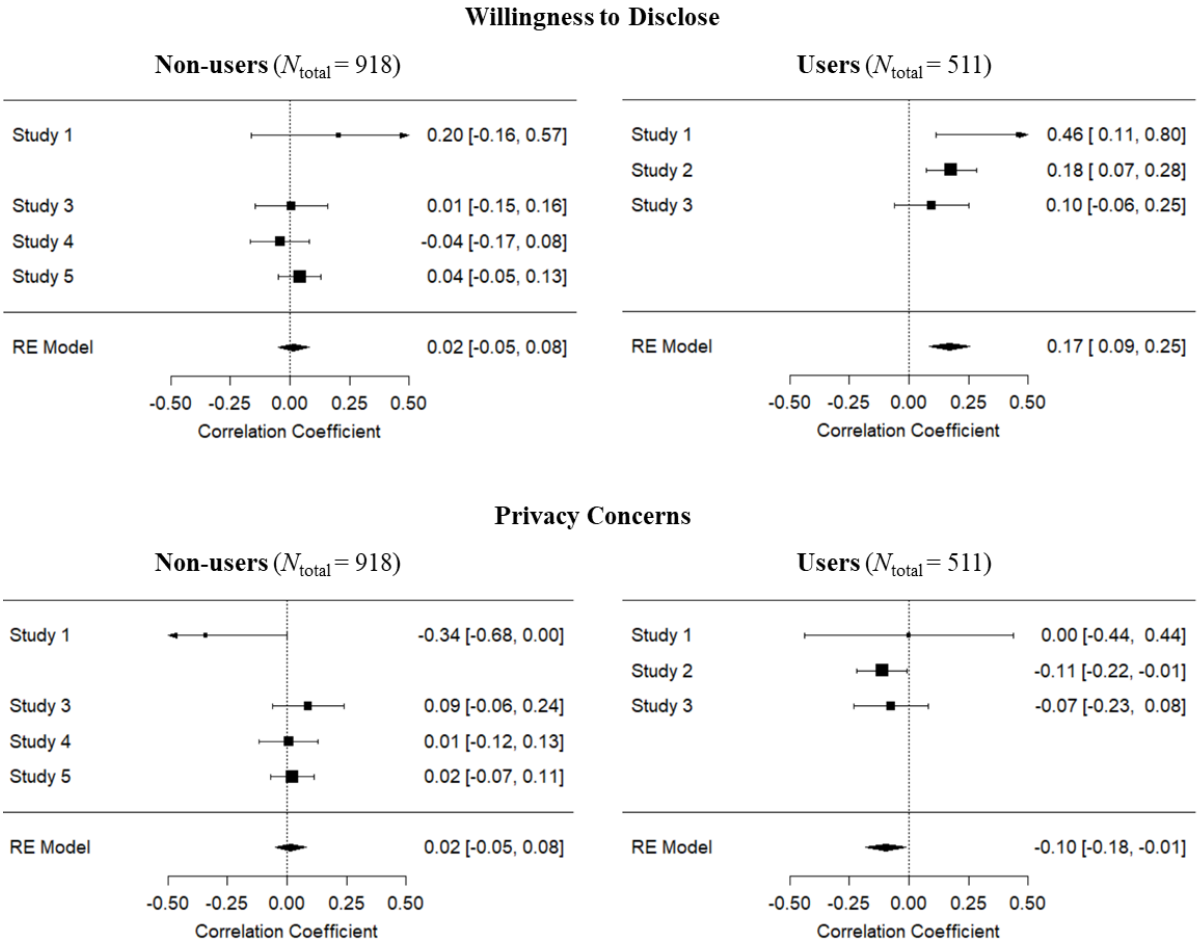
We conducted meta-analyses across all data to test the robustness of the effects and check for potential differences between non-users and users across studies. For this purpose, we transformed *t*-values of β-regression coefficients in Studies 1.1 and 1.2 as well as *F*-values from simple comparisons in Studies 1.3-1.5 into *r* separately for non-users and users for all studies – including Study 1.1 in which users and non-users were not reported separately above. Due to the small sample size in this study, the respective confidence intervals are rather wide.

We expected that higher intellectual competencies would predict more willingness to disclose information (H1a). For non-users, this prediction was not supported, $r = .02$, $z = 0.51$, $p = .609$ (Figure 4, upper left). In contrast, users were more willing to disclose information if perceiving higher intellectual competencies, $r = .17$, $z = 3.98$, $p < .001$ (Figure 4, upper right). Of note, as this effect was only significant in Studies 1.1 and 1.2 but not in Study 1.3, experimental support for H1a is also lacking for the user subsample.

We further expected that higher intellectual competencies would predict less privacy concerns (H1b). This hypothesis was likewise not supported for non-users, $r = .02$, $z = 0.52$, $p = .605$ (Figure 4, bottom left), but for users, $r = -.10$, $z = -2.21$, $p = .027$ (Figure 4, bottom right).

Figure 4

Results of Meta-Analyses Across Studies 1.1-1.5 for the Effect of Intellectual Competencies on Willingness to Disclose (H1a) and Privacy Concerns (H1b) in the Non-User (N = 918) and User (N = 511) Samples

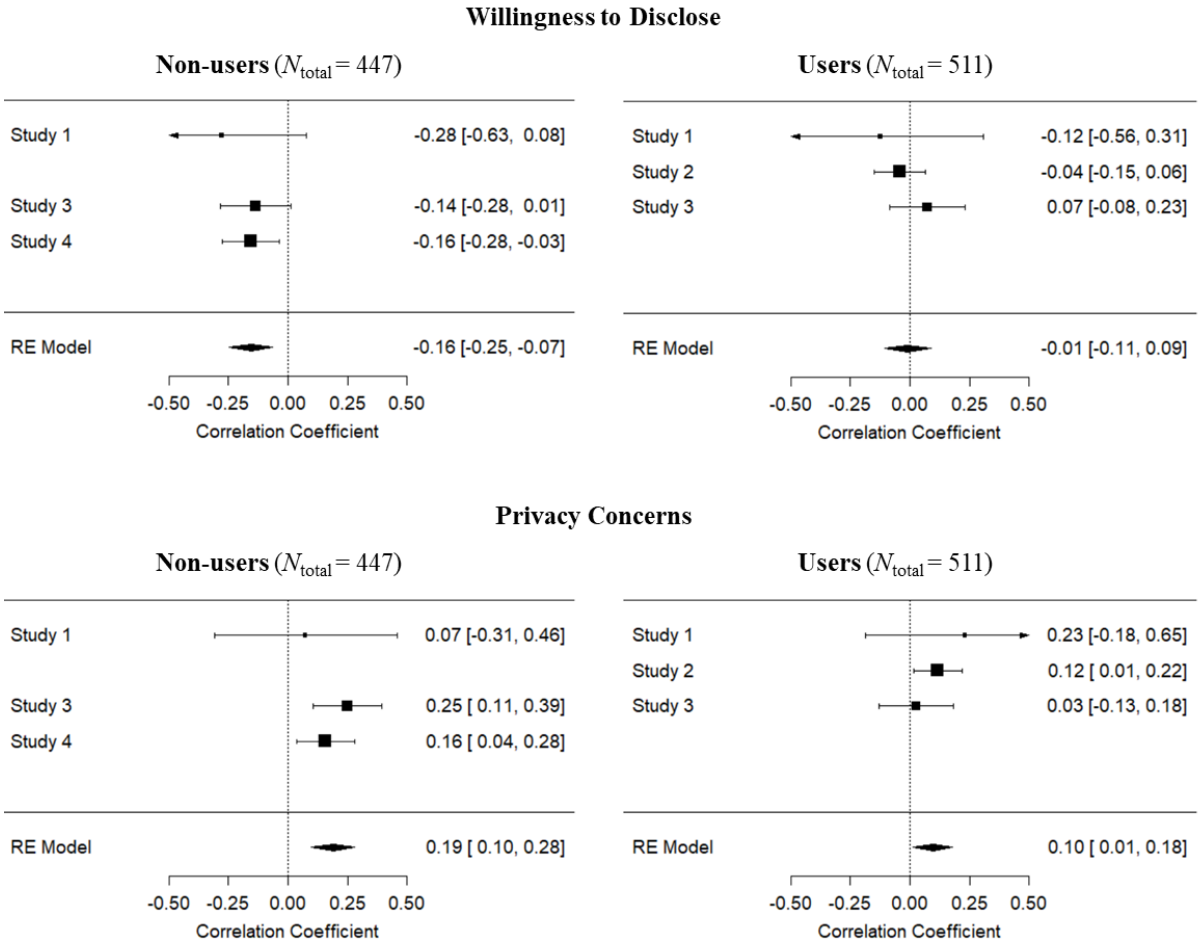


Moreover, we assumed that higher meta-cognitive heuristics would predict less willingness to disclose (H2a). This hypothesis was supported across studies for non-users, r = -.16, z = -3.37, p = .001, but not for users, r = -.01, z = -0.18, p = .858 (Figure 5, top).

Lastly, we assumed that higher meta-cognitive heuristics would predict higher privacy concerns (H2b). As shown in Figure 5 (bottom), this effect was found for non-user and users. It is noteworthy, that the estimated overall effect size was stronger for non-users, r = 0.19, z = 4.13, p < .001, than for users, r = 0.10, z = 2.23, p = .026.

Figure 5

Results of Meta-Analyses Across Studies 1.1-1.5 for the Effect of Meta-Cognitive Heuristics on Willingness to Disclose (H2a) and Privacy Concerns (H2b) in the Non-User (N = 447) and User (N = 511) Samples



To sum up, the meta-analytical results suggest that *intellectual competencies* of a conversational AI led to a higher willingness to disclose and lower privacy concerns for users, whereas no effect of these competencies was observable for non-users. *Meta-cognitive heuristics* in a conversational AI led to more privacy concerns for non-users and users, however, only for non-users they also led to a lower willingness to disclose, whereas, for users, this willingness was not affected by meta-cognitive heuristics.

Discussion of Chapter 2

The present research tested whether perceived competencies of conversational AI affect people's disclosure in terms of privacy concerns and willingness to disclose. Focusing on participants' subjective perception of human-like competencies by applying the action regulation theory (Hacker & Sachse, 2014; Semmer & Frese, 1985), we followed the idea that

technologies are perceived and treated as social actors (Reeves & Nass, 1996) and are ascribed human-like characteristics (Epley et al., 2007).

We predicted that different competencies of conversational AI would have different effects on disclosure. In detail, higher intellectual competencies in a conversational AI should heighten the willingness to disclose (H1a) and lessen privacy concerns (H1b), whereas meta-cognitive heuristics should heighten privacy concerns (H2a) and lessen willingness to disclose (H2b). As differences between users and non-users became evident within the research process and are highly plausible (as discussed below), we differentiated between those two groups.

Meta-Cognitive Heuristics

As expected, meta-cognitive heuristics reduced non-users' willingness to disclose (although the effect was significant only in Study 1.4, the overall pattern of observed effects clearly supports this notion) and heightened their privacy concerns. These findings support previous research suggesting that technical systems can be perceived as too competent – especially when it comes to human-like competencies. Adding to previous literature, we observed that too high competencies do not only lead to feelings of eeriness and perceptions of threat (Stein et al., 2019, 2020; Złotowski et al., 2017) but also to concrete concerns about privacy and a reduced willingness to disclose in the specific case of conversational AI.

However, meta-cognitive heuristics did not affect users' willingness to disclose. Across studies, meta-cognitive heuristics were associated with heightened privacy concerns for users. Given that this relation was significant only in the correlational Study 1.2, it remains unclear whether there is a causal effect. In sum, meta-cognitive heuristics have, if at all, a small effect on users' privacy concerns. It is plausible that for users (in comparison to non-users), benefits due to a conversational AI's heightened competencies on the level of meta-cognitive heuristics are again more obvious and salient and, thus, also more likely to outweigh the perceived negative aspects of information sharing (e.g., higher privacy concerns) in a privacy trade-off. As previous research has further shown a gap between disclosure intentions and actual disclosing behavior (e.g., Norberg et al., 2007), the differences between privacy concerns and actual disclosure might be even larger than the ones observed in our studies focusing on willingness to disclose.

Besides, the findings regarding meta-cognitive heuristics are consistent with another mechanism: high human-like competencies in a technical system might be concerning in the first place (as demonstrated for the non-users) but with ongoing usage, people might accustom to these high capabilities and, thus, be less concerned about them. With the ongoing technological progress, especially in the field of AI, this mechanism raises awareness that particularly

users should be sensitized for potential privacy risks associated with the usage of technical systems as they might lose caution over time.

Intellectual Competencies

For non-users, intellectual competencies of conversational AI did neither affect their willingness to disclose nor their privacy concerns. For users, the results suggest that intellectual competencies heighten their willingness to disclose and lessen their privacy concerns, although we could not validate these effects in an experimental setting (Study 1.3) and, thus, the causality of the effects is not ensured by our data.

The finding that intellectual competencies in a conversational AI are positively related to users' disclosure is in line with prior research showing that people are willing to share their personal data in exchange for benefits. Further, the finding adds to the literature of positive effects of higher (perceived) competencies on technology acceptance (e.g., Dehghani, 2018; McLean & Osei-Frimpong, 2019; Pitardi & Marriott, 2021) by demonstrating a similar effect on users' disclosure. It is, nonetheless, noteworthy that we did not observe the expected effect for non-users of conversational AI. It might be the case that the benefits of intellectual competencies were less clear and salient to non-users and did, thus, not have the same effect as for users. At this point, this is however speculative and should be tested in further research.

Strengths and Limitations

Our findings are based on large samples from several studies (including both survey and experimental data) and allow insight on both users' and non-users' disclosure towards conversational AI. The results clearly signalize that the proposed systematic differentiation of competencies was informative to disentangle the consequences of perceived human-like competencies on disclosure towards conversational AI. Thus, the proposed systematic differentiation of perceived competencies can be a starting point for future research focusing, for instance, on the effects of these competence levels on other aspects of acceptance (e.g., trust, believability) as well as usage (e.g., adoption, usage frequency, usage purposes) for conversational AI or other highly competent (AI) technologies.

Despite the aforementioned strengths of our studies, some limitations should be considered when interpreting the results. Our experimental studies were based on vignette descriptions rather than real interaction with the conversational AI. This approach comes close to capturing responses to information about technology updates. However, as users already have the regular experience of interacting with conversational AI, it was considerably harder to delude them about the system's competencies by using vignette descriptions. Here, the survey data

gave valuable insights especially for the users: while their interaction experience with conversational AI potentially reduced the effectiveness of experimental manipulations, it also enabled them to answer survey question more validly (e.g., they probably have a more realistic estimate of their information sharing). Besides, the experimental manipulation of conversational AI competencies in a real interaction is challenging: to enable a system to make use of meta-cognitive heuristics (e.g., developing general strategies based on previous interactions), regular interaction over a longer period of time would be necessary. We thus encourage longitudinal research in order to investigate how the effects of perceived competencies develop and change within a longer usage period.

Further, we relied on self-reported disclosure intentions but did not assess actual disclosing behavior. Despite the fact that intentions are generally good predictors of behavior, future research should, thus, complement our findings by applying the proposed framework to investigate actual disclosing behavior.

Conclusion

Building on the idea that technologies are anthropomorphized, perceived, and treated as social actors (Epley et al., 2007; Reeves & Nass, 1996), the present research investigated the effects of perceived human-like competencies in a conversational AI on users' and non-users' disclosure. In a nutshell, the results indicate that non-users are concerned about their privacy and do not want to share personal information with highly competent conversational AI (i.e., using meta-cognitive heuristics). In contrast, users are willing to share personal information in exchange for intellectual competencies of the technology. Even though meta-cognitive heuristics in a conversational AI slightly heightened privacy concerns also for users, these competencies did not reduce their willingness to share information with the system. Thus, the presented findings highlight a tension: on the one hand, users can benefit from higher competencies of a conversational AI and are willing to share personal data in exchange. On the other hand, high competencies are perceived as concerning – especially for people with little interaction experience with the system.

The following chapter contains a manuscript that is the result of a cooperation between Miriam Gieselmann (first author) and Prof. Dr. Kai Sassenberg (second author). The following table depicts the contributions of the PhD candidate (and of the co-author, respectively) to the manuscript as well as the current status in the publication process:

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
Miriam Gieselmann	1	70	80	80	80
Kai Sassenberg	2	30	20	20	20
Title of Paper:		The relevance of perceived interactivity for disclosure towards conversational artificial intelligence			
Status in publication process		Content identical version published as: Gieselmann, M., & Sassenberg, K. (2023). The relevance of perceived interactivity for disclosure towards conversational artificial intelligence. In H. Degen & S. Ntoa (Eds.), *Artificial Intelligence in HCI. HCII 2023. Lecture Notes in Computer Science* (Vol. 14051, pp. 55-67). Springer. https://dx.doi.org/10.1007/978-3-031-35894-4_4			

Chapter 3: Characteristics of the Interaction

Nowadays, artificial intelligence (AI) becomes more and more present in our everyday life. One prominent example is conversational AI (also referred to as voice assistants, virtual intelligent agents, etc.): systems like Google Assistant or Amazon Alexa are part of many people's everyday life. Being used, for example, for information access, entertainment, online shopping, or the control of smart home devices, 3.25 billion of these systems had been in use worldwide in 2019 (Statista, 2022). By now, 35% of US adults own a smart speaker with a conversational AI (Kinsella, 2021).

Conversational AI offers convenience, utility, enjoyment, and personalization to its users, but at the same time, gathers a lot of personal information about them (e.g., information about their interests, location data, or access to other applications). While this information might be necessary for several functionalities (e.g., for personalized recommendations or location-based requests), it still raises the issue of privacy concerns regarding the amount of personal data collected by conversational AI (Dubiel et al., 2018; Liao et al., 2019).

As people nonetheless share personal data with conversational AI, it is important to understand which factors determine disclosure (i.e., privacy concerns and willingness to

disclose). Previous research has often focused on determinants such as trust in the provider of the technology (Joinson et al., 2010; Pal et al., 2020), perceptions of the technology itself (e.g., competence; Gieselmann & Sassenberg, 2023b) or rather objective system characteristics (e.g., anthropomorphic design features, Ha et al., 2021; Lucas et al., 2014). Another important aspect that seems to heighten disclosure in several contexts is interactivity (Martelaro et al., 2016; Walsh et al., 2020). However, the conceptualization and operationalization of interactivity often vary between studies.

The present research thus aimed for a systematic approach to investigate the role of different aspects of interactivity for disclosure towards conversational AI at the same time. To this end, we followed the approach of Liu (2003), who differentiates three facets of interactivity: (1) *active control*, (2) *reciprocal interaction* (originally called two-way interaction), and (3) *synchronicity*.

While it has been shown that each of these dimensions can be relevant for disclosure (Green et al., 2016; Y. W. Park & Lee, 2019; Schouten et al., 2007; Walsh et al., 2020), they have barely been investigated at the same time. As research on the effects of interactivity on disclosure towards conversational AI is lacking, the present research investigates the associations between the three afore-mentioned dimensions of interactivity and disclosure towards conversational AI.

Interactivity

Interactivity refers to the experience of responsiveness from another entity, which can be both a person or a technology (Lew et al., 2018), resulting in an exchange between interaction partners. While the concept of interactivity originally stems from interactions between humans, it has become more and more important in technology interactions as well. Interactivity is a core evaluation criterion for technologies and has received considerable attention in research on technology interactions as it enables active, efficient, and quick interactions between users and technologies. Accordingly, previous research has repeatedly investigated the role of interactivity for several aspects of acceptance, such as credibility (e.g., Johnson & Kaye, 2016; Sundar, 2008), satisfaction (e.g., Shu, 2014), usage intentions (e.g., Shin et al., 2013), and disclosure.

Despite the large amount of research already conducted in the field of interactivity, there has been little agreement on how interactivity should be conceptualized – resulting in different conceptualizations across studies (Heeter, 2000; Shin et al., 2013). One promising approach to assess interactivity has been proposed by Liu (2003) distinguishing the three dimensions active control, reciprocal interaction (originally called two-way interaction), and

synchronicity. While there has been some research taking all three dimensions into account and showing positive relations with different acceptance measures (e.g., Fan et al., 2017; Cheng, 2014), such studies are lacking for the relationship between interactivity and disclosure. Nonetheless, some studies focused on one of these (interrelated) dimensions to understand acceptance in general and disclosure in specific.

Active control

Active control refers to the user's perception in how far they have control over the interaction and the exchanged information (Y. Liu, 2003) and can be related to different aspects of acceptance. For a related construct, namely perceived behavioral control, positive relationships with usage intention and ease of use have been observed for social robots (de Graaf & Ben Allouch, 2013). Further, autonomous and thus, less controllable robots, are perceived as more threatening to humans and evoke stronger negative attitude towards robots in general and more opposition to robotics research than non-autonomous robots (Złotowski et al., 2017).

More important to the present research, it has also been observed that controllability was related to more disinhibition, which in turn predicted more online self-disclosure in instant messaging and social media (Green et al., 2016; Schouten et al., 2007). Therefore, it seems reasonable to assume that active control predicts more disclosure.

Reciprocal Interaction

Reciprocal interaction (originally: two-way interaction; Y. Liu, 2003), describes a two-way flow of information. Previous research has often focused on technologies as a medium to enable a two-way flow of information between two humans or between companies and their customers (i.e., computer-mediated communication). For example, anticipated reciprocity in micro blogging predicted higher perceived gratification which in turn predicted more content contribution (X. Liu et al., 2020) and perceived responsiveness is associated with more disclosure on social networks (Walsh et al., 2020). Besides, in the context of companies' corporate social responsibility communication, interactivity (in terms of replies and references to earlier messages) led to greater perceived contingency (i.e., a higher degree to which later messages refer to earlier messages) which in turn led to greater willingness to comment on social media (Lew & Stohl, 2022).

However, with more recent technologies such as social robots or conversational AI, users do no longer interact with another human but with the technology itself. Thus, the focus lays on the two-way flow of interaction between the user and the technology rather than between the user and another human. While research on this type of two-way interaction is scarce,

it has been observed that, in a learning setting, participants reported more disclosing to an expressive robot (e.g., robot changing color, moving arms or head) compared to a non-expressive robot (Martelaro et al., 2016). Hence, reciprocal interaction might likewise be associated with disclosure.

Synchronicity

Synchronicity refers to the perception that the exchange of information between communication partners happens quickly (Y. Liu, 2003). In face-to-face interactions between humans, conversations with multiple (compared to fewer) time lapses lead to lower ratings of the interaction partner's communicative competency (McLaughlin & Cody, 1982). Besides, also in contexts of computer-mediated communication, significant pauses between comments can be associated with negative consequences, such as less trust among communication partners (Kalman et al., 2010). Further, customer service agents responses without delay (vs longer delays) lead to higher perceived co-presence and service quality (E. K. Park & Sundar, 2015).

Focusing more specifically on disclosure, it has been proposed that asynchronicity of communication in interactions between humans could enhance disclosure as there will not be an immediate reaction of the interaction partner (Suler, 2012). However, for personal use contexts of mobile messengers, it has been observed that immediacy of feedback led to more intimacy (Y. W. Park & Lee, 2019). In sum, results do not allow a conclusion regarding the relation between synchronycity and disclosure.

Differentiation of Disclosure

Previous research has shown that people do not necessarily stop sharing their personal data when having privacy concerns (referred to as the "privacy paradox", for a review see Kokolakis, 2017). Thus, we differentiated between the behavioral intention to share personal data with a conversational AI, i.e., the *willingness to disclose*, and more abstract attitudes, i.e., *privacy concerns*. Accordingly, we aimed to investigate:

RQ: How is perceived interactivity (i.e., active control, reciprocal interaction, and synchronicity) of conversational AI associated with disclosure (i.e., willingness to disclose and privacy concerns)?

Current Research

To answer this question, we conducted two online surveys. In the first study, we exploratory investigated the association between the three interactivity dimensions and disclosure

towards conversational AI in a small sample. In the second study, we aimed to replicate our findings in a larger sample of actual users of conversational AI.

Study 2.1

Method

The studies reported here were conducted within a larger research project focusing on another research question than the one discussed here and published in a paper by Gieselmann and Sassenberg (2023b). All results reported in the present paper have not been included in the previous paper and vice versa. Data (Gieselmann & Sassenberg, 2022, Studies 1 and 2) and code for analyses (Gieselmann & Sassenberg, 2023a) are available on PsychArchives.

Design and Participants. Participants were recruited via Prolific for a 10-min online survey renumerated with £1.25. In order to reach stable correlations, we aimed for 300 valid cases (Schönbrodt & Perugini, 2013). Participants were randomly assigned to one of five experimental conditions in which they were surveyed about the perception of a specific technology (i.e., conversational AI, search engine, streaming service, autonomous vehicle, washing machine). After exclusion as pre-registered (https://aspredicted.org/tq3m3.pdf), the sample consisted of $N = 358$ participants (59.8% male, 38.5% female, 1.4% non-binary; aged: $M = 25.5$, 18-62 years). Of these participants, $n = 72$ focused on the perception of conversational AI (54.2% male, 44.4% female, 1.4% non-binary; aged: $M = 27.0$, 18-56 years).

Procedure. After having been randomly assigned to one of the five experimental conditions, participants read a short description of a conversational AI (or another technology – depending on the experimental condition as described above) and were afterwards surveyed about their perceptions of the technology's interactivity as well as their disclosure towards the respective technology. A complete list of all measures taken, instructions, and items is provided in the supplement of Gieselmann and Sassenberg (2023b).

Measures. For the assessment of interactivity we adjusted and augmented the items of (Y. Liu, 2003) for active control, reciprocal interaction, and synchronicity. Based on an exploratory factor analysis (see Table 3), we excluded items 10 and 11 intended for the reciprocal interaction scale because they were loading highest on another factor. Due to the results of Study 2.2 (reported below), we further excluded item 5 from the active control scale to have an identical scale across studies. Accordingly, interactivity was assessed distinguishing *active control* (5 items), *reciprocal interaction* (4 items), and *synchronicity* (6 items).

Table 3

Results From a Factor Analysis for Interactivity (Study 2.1: N = 358)

Item		Factor		
		1	2	3
1	I can choose freely what ... does.		**0.77**	
2	I have little influence on ...'s behavior. [a]		**0.47**	-0.36
3	I can completely control		**0.77**	
4	While interacting with ..., I have absolutely no control over what happens. [a]	0.22	**0.50**	-0.38
5	My actions determine ...'s behavior. [b]		**0.51**	
6	I have a lot of control over my interaction experience with	0.25	**0.69**	
7	... facilitates two-way communication between itself and its user.	0.20		**0.65**
8	... makes me feel like it wants to interact with me.			**0.77**
9	Using ... is very interactive.	0.22		**0.60**
10	... is completely apathetic. [ab]		-0.27	0.21
11	... always reacts to my requests. [b]	**0.50**	**0.43**	
12	Both, ... and I, can start interactions with one another.			**0.74**
13	... processes new input quickly.	**0.73**		
14	...'s reactions come without delay.	**0.71**		
15	... responds very slow. [a]	**0.66**		-0.28
16	I never have to wait for ...'s output.	**0.48**		0.38
17	... immediately answers to requests.	**0.71**	0.23	
18	When I interact with ..., I get instantaneous feedback.	**0.65**		0.34

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings < .2 are not displayed. Factor loadings above .40 are in bold. '…' was replaced with a specific technology depending on the experimental condition (i.e., the voice assistant, the washing machine, the search engine, the streaming service, or the autonomous vehicle).

[a] Item reverse coded. [b] Item not included in final scale.

Disclosure was assessed as willingness to disclose (5 items, e.g., 'I would provide a lot of information to the voice assistant about things that represent me personally'; partially adapted from Gerlach et al., 2015) and privacy concerns (4 items, e.g., 'I feel that the voice assistant's practices are an invasion of privacy'; adapted from Alge et al., 2006). All scales were

answered on seven-point scales (1 = *strongly disagree,* 7 = *strongly agree*). Means, standard deviations and reliabilities for all scales are reported in Table 4.

Table 4

Scale Reliabilities, Means and Standard Deviations for Interactivity and Disclosure (Study 2.1: n = 72)

	Number of items	α	*M*	*SD*
Active control	5	0.61	4.48	0.90
Reciprocal interaction	4	0.51	4.25	0.99
Synchronicity	6	0.73	4.64	0.80
Willingness to disclose	4	0.80	4.49	1.25
Privacy concerns	5	0.90	3.60	1.46

Results

Bivariate correlations (see Table 5) show that all three dimensions of interactivity were positively and significantly correlated with willingness to disclose. Besides, active control and synchronicity were negatively and significantly correlated with privacy concerns, whereas the negative relation between reciprocal interaction and privacy concerns was non-significant.

Table 5

Bivariate Correlations Between Interactivity and Disclosure (Study 2.1: n = 72)

	1	2	3	4
1 Active control				
2 Reciprocal interaction	0.01			
3 Synchronicity	0.28*	0.18		
4 Willingness to disclose	0.24*	0.49***	0.33**	
5 Privacy concerns	-0.27*	-0.23	-0.27*	-0.64***

We further tested the relative importance of the three interactivity dimensions for (1) willingness to disclose and (2) privacy concerns by conducting two separate multiple regression analyses including the three interactivity dimensions as predictors at the same time. Willingness to disclose was associated with reciprocal interaction, $\beta = 0.46$, $p < .001$, 95%-CI[0.25, 0.66], but not with active control, $\beta = 0.17$, $p = .098$, 95%-CI[-0.03, 0.38], nor synchronicity, $\beta = 0.20$, $p = .068$, 95%-CI[-0.01, 0.41]. Further, privacy concerns were not related to neither active control, $\beta = -0.22$, $p = .068$, 95%-CI[-0.45, 0.02], nor reciprocal

interaction, β = -0.20, p = .089, 95%-CI[-0.42, 0.03], nor synchronicity, β = -0.17, p = .159, 95%-CI[-0.41, 0.07].

Discussion

While all interactivity dimensions were correlated to willingness to disclose and privacy concerns (although one correlation was non-significant), the regression analyses showed a more differentiated picture. When controlling for the other two interactivity dimensions in multiple regression analyses, only higher perceived reciprocal interaction was associated with a higher willingness to disclose information towards conversational AI. Apart from that, none of the interactivity dimensions was related to neither willingness to disclose nor privacy concerns.

However, the sample size of this study was rather small and was not limited to users of conversational AI – thus, participants might have lacked sufficient interaction experience with conversational AI. Further, some of the scales had non-satisfactory internal consistencies. To overcome these limitations, we recruited a larger sample of actual conversational AI users in Study 2.2 and slightly adjusted some of the scales.

Study 2.2

Method

Design, Participants, and Procedure. Study 2.2 was a conceptual replication of Study 2.1 with a focus on actual users of conversational AI as participants. For a 12-min online study, we recruited participants owning a home assistant or smart hub and having used a conversational AI before via Prolific. After exclusion as pre-registered (https://aspredicted.org/v87ub.pdf), the sample consisted of N = 334 participants (54.2% male, 44.0% female, 1.8% non-binary; age: M = 30.3, 18-73 years).

Measures. We developed two additional items to replace the items of the reciprocal interaction scale excluded in Study 2.1. Exploratory factor analysis (see Table 6) supported the three-factor structure of the interactivity items with the exception of one item intended for the active control scale loading highest on another factor. The respective item was excluded from the scale in this study as well as in Study 2.1. Accordingly, active control and synchronicity were assessed with the same items as in Study 2.1. Reciprocal interaction was assessed using the four items of Study 2.1 and the two additional items 10 and 11.

Table 6

Results From a Factor Analysis for Interactivity (Study 2.2: N = 334)

Item		Factor		
		1	2	3
1	I can choose freely what the voice assistant does.	0.28	0.27	**0.57**
2	I have little influence on the voice assistant's behavior. [a]			**0.73**
3	I can completely control the voice assistant.	0.26		**0.56**
4	While interacting with the voice assistant, I have absolutely no control over what happens. [a]			**0.68**
5	My actions determine the voice assistant's behavior. [b]		0.38	0.32
6	I have a lot of control over my interaction experience with the voice assistant.	0.28	0.33	**0.62**
7	The voice assistant facilitates two-way communication between itself and its user.		**0.67**	0.24
8	The voice assistant makes me feel like it wants to interact with me.		**0.69**	
9	Using the voice assistant is very interactive.	**0.40**	**0.55**	0.21
10	The voice assistant gives me the opportunity to interact with it.		**0.55**	0.31
11	The voice assistant encourages me to interact with it		**0.68**	
12	Both, the voice assistant and I, can start interactions with one another.		**0.66**	
13	The voice assistant processes new input quickly.	**0.57**	0.30	0.25
14	The voice assistant's reactions come without delay.	**0.77**		
15	The voice assistant responds very slow. [a]	**0.69**		
16	I never have to wait for the voice assistant's output.	**0.68**		
17	The voice assistant immediately answers to requests.	**0.79**		
18	When I interact with the voice assistant, I get instantaneous feed-back.	**0.77**	0.20	

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings < .2 are not displayed. Factor loadings above .40 are in bold.

[a] Item reverse coded. [b] Item not included in final scale.

Willingness to disclose was assessed with the same items as in Study 2.1, whereas we added an additional item in the scale for privacy concerns (i.e., 'I am concerned about my privacy when using the voice assistant.'). Means, standard deviations and reliabilities for all scales are reported in Table 7.

Table 7

Scale Reliabilities, Means and Standard Deviations for Interactivity and Disclosure (Study 2.2: N = 334)

	Number of items	α	*M*	*SD*
Active control	5	0.71	4.67	1.01
Reciprocal interaction	6	0.76	4.18	1.05
Synchronicity	6	0.60	4.73	0.95
Willingness to disclose	5	0.90	4.03	1.35
Privacy concerns	5	0.90	3.57	1.46

Results

Bivariate correlations (see Table 8) show that all three interactivity dimensions are significantly positively related to willingness to disclose and negatively related to privacy concerns.

Table 8

Bivariate Correlations Between Interactivity and Disclosure (Study 2.2: N = 334)

	1	2	3	4	5
1 Active control					
2 Reciprocal interaction	0.25***				
3 Synchronicity	0.40***	0.35***			
4 Willingness to disclose	0.11*	0.29***	0.14*		
5 Privacy concerns	-0.13*	-0.13*	-0.16**	-0.50***	

* p < .05. ** p < .01. *** p < .001.

As in Study 2.1, we tested the relative importance of the three dimensions of interactivity for (1) willingness to disclose and (2) privacy concerns by conducting two separate multiple regression analyses (pre-registered): willingness to disclose was associated with reciprocal interaction, β = 0.27, $p < .001$, 95%-CI[0.16, 0.38], but not with active control, β = 0.03, $p = .654$, 95%-CI[-0.09, 0.14], nor synchronicity, β = 0.03, $p = .571$, 95%-CI[-0.08, 0.15]. Further, privacy concerns were not related to neither active control, β = -0.07, $p = .248$,

95%-CI[-0.19, 0.05], nor reciprocal interaction, β = -0.07, *p* = .224, 95%-CI[-0.19, 0.04], nor synchronicity, β = -0.11, *p* = .073, 95%-CI[-0.23, 0.01].

Discussion

The results replicate the findings of Study 2.1. Active control, reciprocal interaction, and synchronicity are all correlated positively with willingness to disclose and negatively with privacy concerns. As in Study 2.1, when controlling for the influence of the other two interactivity dimensions by using multiple regression analyses, higher perceived reciprocal interaction was again associated with a higher willingness to disclose information towards conversational AI. Beyond this relationship, none of the interactivity dimensions was related to neither willingness to disclose nor privacy concerns.

Discussion of Chapter 3

The research presented in this paper tested the relationship between three dimensions of perceived interactivity (active control, reciprocal interaction, and synchronicity) and disclosure towards conversational AI. The findings support findings from earlier research by showing that active control, reciprocal interaction, and synchronicity are all positively correlated with willingness to disclose and negatively correlated with privacy concerns.

However, the regression analyses revealed that in contrast to active control and synchronicity, reciprocal interaction is the most relevant interactivity dimension to understand people's willingness to disclose personal information towards conversational AI. This finding indicates that the feeling of a two-way flow of information between a user and a conversational AI is crucial for the users' willingness to disclose personal information to the system.

Although we cannot rule out that active control and synchronicity facilitate the perception of two-way interaction, the latter dimension seems to be the most important one for people's decision whether to disclose personal information towards conversational AI.

This implies that the feeling of reciprocal interaction with a conversational AI could heighten users' willingness to share personal information with such a technology. Accordingly, it should be considered that for such highly interactive technologies, special care needs to be taken in order to prevent invasions of the users' privacy.

Strengths and Limitations

Our findings are based on two studies showing very similar result patterns. Thus, the results indicate that the differentiation of interactivity facets as proposed by Liu (2003) is useful to disentangle effects of interactivity on disclosure towards conversational AI. Accordingly,

this conceptualization of interactivity can be a starting point for future research focusing on other aspects of acceptance (e.g., trust, believability, or adoption) of conversational AI or other technologies.

Nonetheless, when interpreting the results, it should be taken into account that this research was based on survey data, thus, not allowing conclusions about causality. Therefore, we highly encourage future research using experimental manipulations of the three interactivity dimensions as well as longitudinal studies focusing, for example, on actual disclosing behavior, to complement our findings – potentially also for other technologies than conversational AI.

Conclusion

In contrast to earlier studies, we assessed active control, reciprocal interaction, and synchronicity to focus on three dimensions of interactivity at the same time. Taken together, our research suggests that interactivity in terms of reciprocal interaction (but not active control or synchronicity) is relevant to understand willingness to disclose towards conversational AI while we observed no significant relationships between interactivity and privacy concerns. In doing so, our findings suggest that reciprocal interaction is (compared to active control and synchronicity) the most relevant interactivity facet to understand peoples' willingness to disclose towards conversational AI.

The following chapter contains a manuscript that is the result of a cooperation between Miriam Gieselmann (first author), Josephine Hagedorn (second author) and Prof. Dr. Kai Sassenberg (third author). The following table depicts the contributions of the PhD candidate (and of the co-authors, respectively) to the manuscript as well as the current status in the publication process:

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
Miriam Gieselmann	1	70	70	80	80
Josephine Hagedorn	2	0	20	0	0
Kai Sassenberg	3	30	10	20	20
Title of Paper:		Do perceived benefits compensate for low provider trustworthiness in disclosure decisions? An experimental investigation			
Status in publication process		Revised version submitted for publication			

Chapter 4: Characteristics of the Interaction and the Interaction Partner

Every day, people receive plenty of algorithm-based recommendations – for example, for social media content (i.e., by a personalized social media feed), products they might want to purchase (in online shopping), movies, podcasts, music, or cooking recipes. In order to give users appropriate recommendations, algorithms need personal information about these individuals. The more information the recommendation system has, the more individualized recommendations it can provide. A guide giving personalized recipe recommendations, for instance, will be better able to provide fitting recipe recommendations, the more information about eating habits and preferences, allergies, intolerances, health status (e.g., overweight), and grocery budget it has been able to collect and incorporate.

People are willing to disclose such information if they perceive to get benefits in exchange (e.g., Barth et al., 2019; Sharma & Crossler, 2014; Wottrich et al., 2018), for example, better personalization, access to useful functions, or enjoyment – or: an output of the system from which they can benefit. Such benefits have further been found to outweigh the perceived risks of disclosure in a privacy calculus (e.g., Joinson et al., 2010; Taddei & Contena, 2013; Wottrich et al., 2018), thus offering an explanation for a phenomenon previously referred to as the privacy paradox, namely the disclosure of personal information despite being concerned about privacy (for a review, see Kokolakis, 2017).

Besides perceived benefits, another important determinant of disclosure is whether one trusts the entity to which the information is disclosed (e.g., another human; Wheeless & Grotz, 1977) – so, in the case of recommendation algorithms: the provider of the respective algorithm (similar to the providers of social networks; Taddei & Contena, 2013; or e-commerce; Schoenbachler & Gordon, 2002; Wakefield, 2013). Thus, the question arises: How do perceived benefits and provider trustworthiness play together? While the interplay of perceived benefits and perceived risks has been repeatedly studied within research on the privacy calculus (e.g., Joinson et al., 2010; Taddei & Contena, 2013), few studies focused on a potential interaction between perceived benefits and provider trustworthiness (for an exception, see Wottrich et al., 2017).

Furthermore, previous research on disclosure has often been limited to survey studies (Dinev & Hart, 2006; Taddei & Contena, 2013) or assessing actual disclosure in the field (e.g., on social media; cf. dataset 2 in Kezer et al., 2022), which do not allow for causal conclusions. The few experimental studies in this field typically rely on hypothetical scenarios (e.g., Beresford et al., 2012, letting participants decide which of two mobile apps they would prefer to download – the cheaper one or, the less intrusive one), whereas experimental studies with actual technology interactions are scarce (for an exception, see Wottrich et al., 2017).

Overall, there is a deficit of research studying (a) the interplay of perceived benefits and provider trustworthiness (i.e., statistically testing the interaction of these constructs) and (b) people's in situ response to the use of a recommendation algorithm allowing to capture the effect of the experience of benefits. Therefore, the present research investigates the potential interplay of perceived benefits (i.e., perceived output quality) and provider trustworthiness for the willingness to disclose in an experimental setting where participants interacted with an actual technology (i.e., a recipe guide giving algorithm-based recommendations). In doing so, we aim to overcome the limitations of survey studies (i.e., not allowing causal conclusions) as well as existing experimental investigations (i.e., participants lacking actual usage experience with the technology) and add to the literature on the relevance and the interplay of perceived benefits and provider trustworthiness for people's willingness to disclose personal information.

People Disclose Information for Benefits

In general, perceived benefits are crucial for people's interaction with technology – for example, perceived usefulness is one of the critical determinants of technology acceptance (cf. Venkatesh & Bala, 2008). Beyond general acceptance of technologies, also people's decision to disclose personal information largely depends on whether they (perceive to) get benefits in exchange for their disclosure. Importantly, perceived benefits do often even outweigh perceived

risks (e.g., privacy concerns) in a privacy calculus (e.g., Barth et al., 2019; Bol et al., 2018; Dienlin & Metzger, 2016; Kezer et al., 2022; Wottrich et al., 2018), stressing the important role of perceived benefits in disclosure decisions.

These benefits can be of various types, including, for example, monetary rewards (e.g., Carrascal et al., 2013) or savings (e.g., Beresford et al., 2012; Yang & Wang, 2009), social benefits (Dienlin & Metzger, 2016; e.g., Kezer et al., 2022; Trepte et al., 2020), enjoyment (e.g., Sharma & Crossler, 2014), or usefulness (e.g., adressed as one of several benefits in the items used by Bol et al., 2018; Sharma & Crossler, 2014), functionality (e.g., Barth et al., 2019), and app value (e.g., Wottrich et al., 2018). The latter ones are more closely related to the benefits investigated in the current research as high perceived output quality should, for instance, be associated with a higher perceived usefulness and a higher perceived app value. Across these various types of benefits, studies have demonstrated that perceived benefits are positively related to disclosure intentions (e.g., Bol et al., 2018; dataset 3 in Kezer et al., 2022; Sharma & Crossler, 2014; Wottrich et al., 2018; Yang & Wang, 2009) and self-reported as well as actual disclosure (e.g., Carrascal et al., 2013; Dienlin & Metzger, 2016; dataset 1 and 2 in Kezer et al., 2022; Trepte et al., 2020).

For example, Sharma and Crossler (2014) conducted an online survey investigating several potential determinants of the behavioral intention to voluntarily disclose information on social media. They observed that, amongst other factors, perceived benefits in terms of enjoyment and usefulness were positively related to disclosure intentions. Furthermore, results of an experimental study by Barth et al. (2019) indicate that perceived benefits in terms of higher functionality, app design, and lower costs were relevant determinants of the decision to download or use a mobile phone application (whereas privacy aspects did not have a significant impact). Accordingly, we expected the following:

H1: Participants share more information, if they perceive the quality of the system output as higher.

(Provider) Trustworthiness Affects Disclosure Intention

Besides perceived benefits, trustworthiness is essential for people's decision to disclose personal information. Trustworthiness refers to the perceptions of others as competent, benevolent, and integer (cf. Mayer et al., 1995). While, at first sight, a low trustworthiness might seem to be strongly related to higher perceived risks (e.g., regarding privacy), it has to be considered that this relationship might differ, for instance, for different dimensions of privacy concerns (cf. Baruh & Cemalcılar, 2014). Furthermore, risks associated with sharing one's data should not be equalized with trustworthiness of the provider, because third parties can

invade trustworthy systems and misuse the data they got access to. Accordingly, trustworthiness and risk perceptions have been treated separately in the literature (e.g., Joinson et al., 2010) and low associations between the constructs have been observed (e.g., Rosenthal et al., 2020). Within the current research, we focus on the perceptions of trustworthiness rather than risk perceptions and concerns. To avoid a confound between trustworthiness and benefits, we aimed at a manipulation of benevolence and integrity, but not of competence – the third component of trustworthiness (Mayer et al., 1995). A more competent provider would be likely to offer a service with higher benefits.

Even in face-to-face interactions between two humans, people consider the other's trustworthiness when deciding whether to share personal information (Wheeless & Grotz, 1977). When people interact via technology, not only the other communicator's trustworthiness is considered but also the trustworthiness of the company providing the technology that mediates the communication: for social networking sites, research has shown that trust in other users (Walrave et al., 2012), as well as trust in the provider of the respective website (Taddei & Contena, 2013), is strongly related to (self-reported) disclosure.

Besides, there are several contexts where people do not interact with another human but solely with a technology provided by a company. For example, in e-commerce, people routinely use websites to shop online. Not surprisingly, also in this context, trust in the provider (i.e., a company) is related to a higher willingness to disclose personal information (Metzger, 2006; Schoenbachler & Gordon, 2002; Wakefield, 2013). Other studies also showed the relevance of provider trustworthiness for the willingness to disclose personal information on health-related websites (Chen et al., 2017) or public e-government services (Beldad et al., 2012), as well as for actual disclosure in web surveys (Joinson et al., 2010) and advergaming (Wottrich et al., 2017). Although some researchers found no relationship between (requestor) trustworthiness and disclosure (intention) (e.g., Norberg et al., 2007), the majority of studies proposes a positive relationship between trustworthiness and (self-reported) disclosure (e.g., Joinson et al., 2010; Taddei & Contena, 2013) as well as disclosure intentions (e.g., Beldad et al., 2012; Chen et al., 2017; Schoenbachler & Gordon, 2002; Wakefield, 2013). Thus, we expected the following:

H2: Participants share less information if the system provider is less (vs. more) trustworthy.

Taken together, both (provider) trustworthiness and perceived benefits can lead people to disclose information. There is initial evidence supporting this notion within the same study: both perceived benefits (e.g., personalization, usefulness) and trust (in an application as well as

its users) have been found to be simultaneously positively related to the willingness to disclose personal information (e.g., Beldad & Kusumadewi, 2015; Pal et al., 2020). But what happens if people interact with a technology that they perceive as more or less beneficial but at the same time do not trust its provider? Can perceived benefits compensate for low provider trustworthiness?

Do Benefits Outweigh low Trustworthiness in Disclosure Decisions?

As described before, when deciding about information disclosure, perceived benefits can outweigh perceived risks (e.g., privacy concerns or low perceived privacy) in a privacy calculus (as proposed by significant interactions between these two constructs, observed, for instance, by Joinson et al., 2010; Taddei & Contena, 2013; or in Study 1 of Wottrich et al., 2018). Differing from those studies, the current research does not consider perceived risks but instead focuses on perceived benefits and provider trustworthiness as determinants of disclosure intentions. It seems reasonable to assume that high perceived benefits could also compensate for low provider trustworthiness (as they do for perceived risks): if people perceive a technical system to provide highly beneficial output (or other benefits), they might be more willing to disclose personal information irrespective of a low trustworthiness of the provider. While some studies have investigated the relevance of perceived benefits and provider trustworthiness simultaneously (e.g., Beldad et al., 2012; Campbell, 2019; Wakefield, 2013), the idea of an interaction between those concepts has barely been tested. As an exception, Wottrich et al. (2017) investigated the interplay of personalization (as a potential benefit), brand trust, and privacy concerns as determinants of information disclosure. Their results indicate an interaction between privacy concerns and brand trust, but no interaction of customization with neither brand trust nor privacy concerns. As their study did further not show a main effect of customization (as perceived benefit) on information disclosure, it is possible that their manipulation of perceived benefits (as customization) was not perceived as beneficial enough to evoke the expected effects. Accordingly, when perceived benefits are stronger, they could still outweigh a low trustworthiness of the provider.

Given that we aimed to study the effects of perceived benefits and provider trustworthiness in the context of app usage – that is, immediately after the experience of the benefits – it seems even more likely that the impact of benefits overrules the one of trustworthiness (compared to a survey in which participants reflect in an abstract manner about both). Thus, we expected that a high perceived benefit (i.e., a system output of higher quality) could reduce the negative effect of low provider trustworthiness on disclosure intention:

H3: The negative effect of low provider trustworthiness will be weaker if participants perceive the quality of the system output as higher.

How are Trustworthiness and Benefits Related to Usage Intention?

While the primary focus of this research was to understand when people are willing to disclose personal information, it is highly plausible that perceived output quality, as well as provider trustworthiness, do affect not only disclosure (intention) but also people's willingness to use or adopt a technology that requires the disclosure of such information (cf. Pal et al., 2020). Although the willingness to disclose and the willingness to use are strongly related constructs, the relevance of perceived benefits (i.e., output quality) and provider trustworthiness for these concepts might differ – when being asked about usage intention, the issue of data disclosure is less salient than in the more tangible decision to disclose specific information. Accordingly, it seems likely that provider trustworthiness could be less important for the intention to use than for the intention to disclose, but perceived benefits (in terms of perceived output quality) should be no less relevant. Thus, we further exploratively investigated:

RQ: How do provider trustworthiness and perceived output quality affect participants' intention to use?

Current Research

We conducted an online study experimentally manipulating provider trustworthiness to test the influence of this factor and its interaction with perceived output quality on people's willingness to disclose. In addition, we exploratively looked at their effect on usage intention. In our study, participants actually interacted with a technology – here: a recipe guide, giving users personalized recipe recommendations based on their input (e.g., information about their eating preferences and allergies). H2 and H3 were pre-registered, while the RQ was exploratively investigated. H1 was not pre-registered, as it is a mere replication of a highly plausible and well-investigated effect but added to facilitate comprehensibility.

Study 3

Method

Design and Participants. We aimed to sample 368 valid cases to achieve a power of 80% and an alpha error of 0.05 for observing a minimum effect of $f^2 = 0.03$ in a multiple regression with three predictors (https://aspredicted.org/Y9F_TKV). We recruited German-speaking participants via the Clickworker panel. Four hundred ninety participants started the study. As preregistered, we excluded participants who failed at least one of two attention checks

(23 cases). Besides we excluded participants based on the following reasons: participating in the study more than once (77 cases), not completing the interaction with the recipe guide (22 cases), technical issues (i.e., not having received five different recipe recommendations or incomplete/missing backend data, 30 cases), or having participated in a previous study with a similar recipe guide (9 cases). All remaining participants matched the pre-registered inclusion criterion of fluently speaking German.

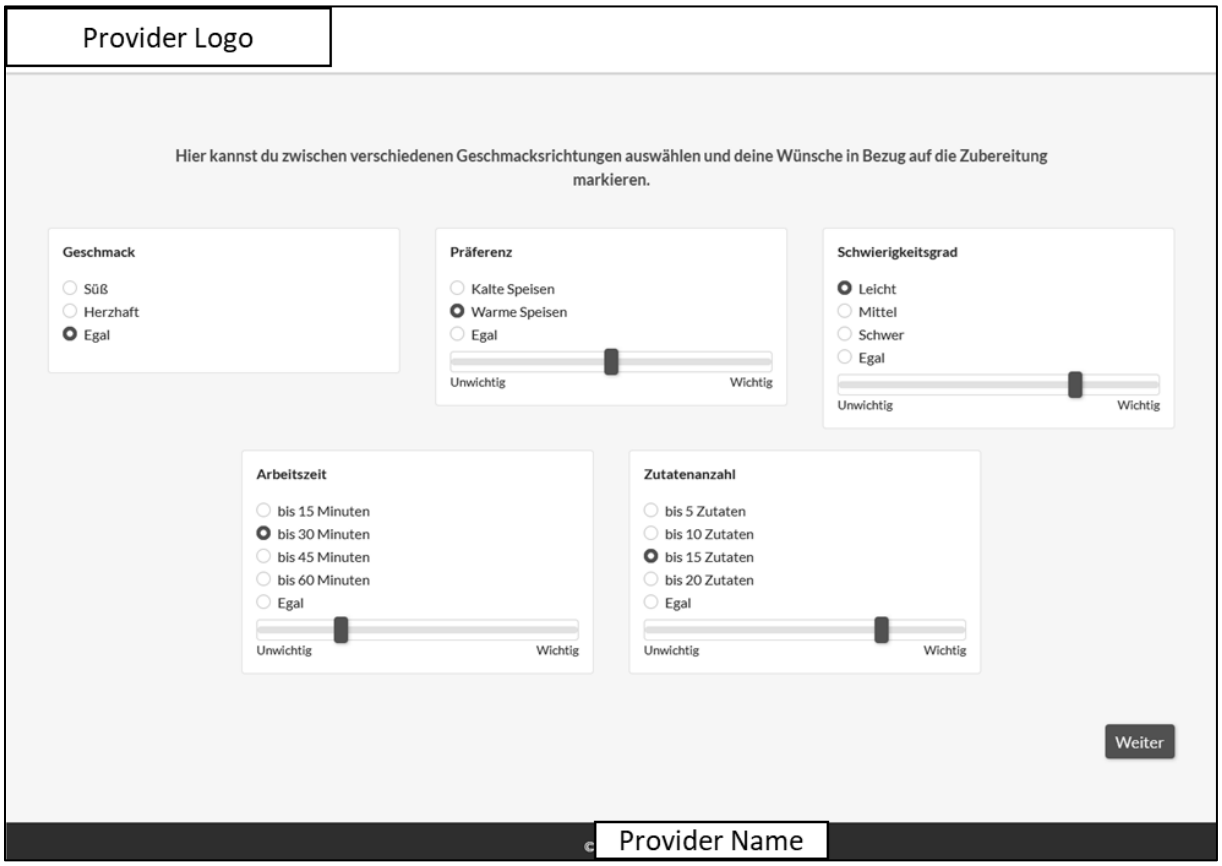
The eligible sample consisted of N = 329 (gender: 36.47% female, 62.92% male, 0.30% non-binary, 0.30% preferred not to say; aged: 18-76, M = 41.54 years). The study was implemented as a one-factorial (provider trustworthiness: low vs. high) between-subjects design.

Procedure. Participants were invited to a study evaluating a newly developed recipe guide. Depending on the randomly assigned experimental condition, participants were informed that the recipe guide was provided on the website of a well-known food company (low trustworthiness) or a well-known health insurance company (high trustworthiness) and received brief information about the respective provider. We chose these providers based on a pre-test with an independent sample of N = 149 participants. Out of eight potential providers of a recipe guide, the food company was perceived as least trustworthy (M = 2.99, SD = 1.30, on a 7-point Likert scale with low values indicating low perceived trustworthiness), and the health insurance company as the most trustworthy (M = 5.04, SD = 1.30), $t(149) = 16.20$, $p < .001$, d = 1.32. Except for this reference to the (ostensible) provider, both conditions were identical.

Participants then interacted with a recipe guide. This recipe guide (as used in Horstmann et al., 2023) was developed at Bielefeld University as part of the IMPACT project (https://www.impact-projekt.de/) and adjusted for the purpose of this study. Based on participants' information on eating restrictions and preferences as well as their importance (see Figure 6 for an example screen asking for participants' input), the recipe guide recommended them five individual breakfast recipes. The recipes were selected by an algorithm using the participants' input. Participants had to click through all five recipes and rate how much they liked each recommendation (1-5 stars, see Figure 7).

Figure 6

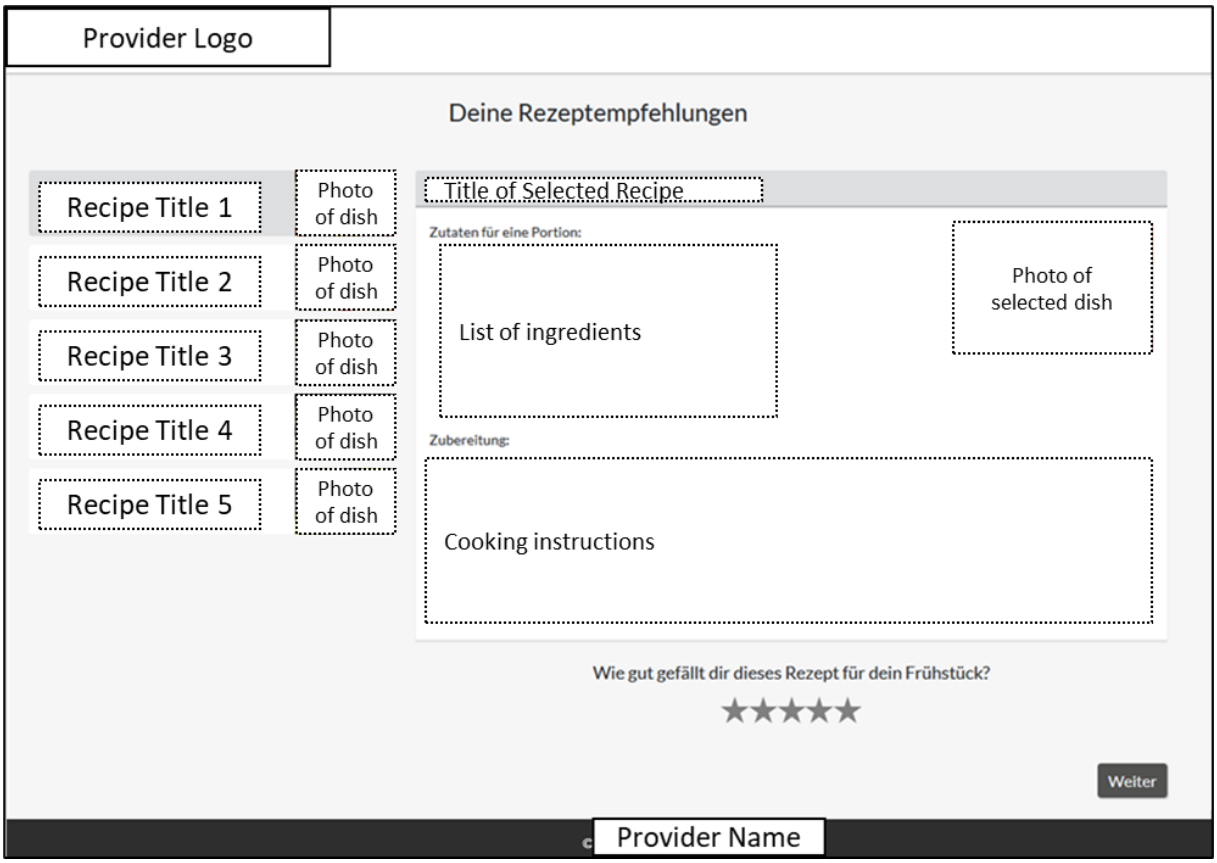
Screenshot of the Recipe Guide Asking for Eating Preferences and Their Importance



Note. Depending on the experimental condition, participants saw different provider logos and names (i.e., food company or health insurance company). Besides, the color of the header was adjusted to fit the color of the respective logo.

Figure 7

Screenshot of the Recipe Guide Showing Recipe Recommendations (Example)



Note. Depending on the experimental condition, participants saw different provider logos and names (i.e., food company or health insurance company). Besides, the color of the header was adjusted to fit the color of the respective logo. Each participant received personalized recommendations based on their input. Due to copyright reasons, recipes and accompanying photos are covered by placeholders.

After completing the interaction with the recipe guide, participants were asked about (1) the perceived quality of the recipe guide's output, (2) their willingness to disclose additional information to the recipe guide to receive better-personalized recommendations, (3) their intention to use the recipe guide, and (4) perceived provider trustworthiness (manipulation check). A complete list of measures taken is provided in Appendix C.

Measures. *Perceived output quality* was measured with four self-developed items ($\alpha = 0.94$, $M = 4.93$, $SD = 1.35$; e.g., 'I am satisfied with the recipe selection,' item translated from German – this applies to all items reported). Apart from willingness to disclose, all items were assessed on a seven-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*).

We measured *willingness to disclose* as the proportion of items participants were willing to share from a list of 20 items (e.g., age, weight, budget for groceries) to receive better-

fitting recipe recommendations (α = 0.86, *M* = 0.42, *SD* = 0.21; for details see Appendix C, Table C1). Participants had to select whether they would share the respective information for each item. According to an exploratory factor analysis, one can further differentiate willingness to disclose information that is (a) strongly related to recipe recommendations (11 items, e.g., budget for groceries, sportive activities; α = 0.86, *M* = 0.65, *SD* = 0.29) or (b) not related to recipe recommendations (7 items, e.g., income, browser history, α = 0.71, *M* = 0.08, *SD* = 0.16); two items (i.e., email and relationship/family status) were not assignable to these factors due to ambiguous factor loadings.

We further assessed *intention to use* (*M* = 4.63, *SD* = 1.49; 'If I had access to the recipe guide, I would use it.') with one item (adjusted from Venkatesh & Bala, 2008). *Provider trustworthiness* (manipulation check) was assessed with six items covering benevolence, integrity, and competence (α = 0.92, *M* = 4.81, *SD* = 1.15; e.g., 'I am convinced that the provider acts in the interest of the users.').

Results

Manipulation Check. As intended and in line with the pretest, perceived provider trustworthiness was lower in the low trustworthiness condition (*M* = 4.50, *SD* = 1.22) compared to the high trustworthiness condition (*M* = 5.16, *SD* = 0.96), $t(322.31) = -5.47$, $p < .001$, $d = 0.60$.

Hypotheses Testing. We expected that willingness to disclose would be higher if the output quality is perceived as higher (H1), lower if the provider is less (vs. more) trustworthy (H2), and that the negative effect of low provider trustworthiness will be weaker when participants perceive the quality of the system output as higher (H3). To test these hypotheses, we conducted a multiple regression analysis (pre-registered) regressing willingness to disclose on perceived output quality, provider trustworthiness, and the interaction of these two factors. Perceived output quality was positively associated with willingness to disclose, *B* = .029, *SE* = .012, β = 0.13, $t(325) = 2.42$, $p = .016$, supporting H1. Higher provider trustworthiness led to a higher willingness to disclose, *B* = .027, *SE* = .012, β = 0.13, $t(325) = 2.32$, $p = .021$, supporting H2. The interaction of the two factors was non-significant, *B* = -.010, *SE* = .012, β = -0.05, $t(325) = -0.82$, $p = .411$, thus, not supporting H3.

Exploratory Analyses. We further observed that the above-described effects held only for the willingness to disclose information related to recipe recommendations but not for unrelated information (see Table 9).

Table 9

Regression of Willingness to Disclose Information (1) Related and (2) Unrelated to Recipe Recommendations on Perceived Output Quality and Provider Trustworthiness (N = 329)

	Willingness to disclose									
	Related information					Unrelated information				
	B	*SE*	β	*t*(325)	*p*	*B*	*SE*	β	*t*(325)	*p*
Output quality	**0.03**	0.02	**0.11**	2.07	.039	0.02	0.01	0.09	1.65	.101
Provider trust-worthiness	**0.04**	0.02	**0.15**	2.76	.006	0.00	0.01	0.03	0.49	.626
Output quality x Provider trust-worthiness	-0.02	0.02	-0.06	-1.05	.293	0.00	0.01	0.01	0.19	.851

Note. Significant coefficients are in bold.

To explore the effects of provider trustworthiness and perceived output quality on the intention to use the recipe guide, we conducted another regression analysis including the intention to use as criterion and perceived output quality, provider trustworthiness, and the interaction of these two factors as predictors. Perceived output quality was positively associated with intention to use, $B = .98$, $SE = .06$, $\beta = 0.66$, $t(325) = 15.54$, $p < .001$. The associations of higher provider trustworthiness, $B = 0.01$, $SE = 0.06$, $\beta = 0.01$, $t(325) = 0.22$, $p = .828$, and the interaction of the two factors, $B = 0.10$, $SE = 0.06$, $\beta = 0.06$, $t(325) = 1.51$, $p = .132$, with intention to use, were non-significant. Thus, only perceived output quality but neither provider trustworthiness nor the interaction of these two factors predicted intention to use the recipe guide.

Discussion of Chapter 4

The present research tested the potential interplay of perceived benefits (in terms of output quality) and provider trustworthiness on disclosure (and usage) intentions right after using an algorithm-based recommendation system. We aimed to overcome shortcomings of existing research (i.e., correlational data and hypothetical scenarios without actual technology interaction) by letting participants interact with an experimentally manipulated recipe guide before asking them about their willingness to disclose personal information to this system.

We predicted that participants would share more information if they perceived the quality of the system output as higher (H1). Besides, we expected them to be less willing to share information when the system provider is less (vs. more) trustworthy (H2) and that this negative effect of low provider trustworthiness would be weaker when participants perceived

the quality of the system output as higher (H3). We further explored the interplay of perceived output quality and provider trustworthiness on the intention to use the recipe guide (RQ).

Willingness to Disclose Information

Perceived output quality and provider trustworthiness were positively related to participants' willingness to disclose personal information (supporting H1 and H2). These findings are in line with previous research showing the relevance of perceived benefits (e.g., Barth et al., 2019; Bol et al., 2018; Carrascal et al., 2013; Sharma & Crossler, 2014; Wottrich et al., 2018; Yang & Wang, 2009) as well as provider trustworthiness (e.g., Joinson et al., 2010; Schoenbachler & Gordon, 2002; Taddei & Contena, 2013; Wakefield, 2013; Wottrich et al., 2017) in disclosure decisions. As effect sizes for both independent variables were similar, their importance for disclosure intentions within our study was comparable. This contrasts an earlier survey study that showed a higher importance of perceived benefits for disclosure than of provider trustworthiness (i.e., by showing a stronger predictive value of benefits than of trustworthiness; Campbell, 2019). Such differences can result from a number of factors such as the specific materials, the wording of the dependent variable or the strength of the applied manipulation. Hence, further research comparing different manipulations and measures for each concept are required to draw ultimate conclusions.

Further, in our study, both effects (i.e., of perceived output quality and provider trustworthiness) were observable for the willingness to share information strongly related to recipe recommendations (e.g., budget for groceries, sportive activities) but not for seemingly unrelated information (e.g., income, location). Thus, the perceived relevance and legitimacy of requested information might be another important factor determining disclosure.

Contrary to our expectations, there was no interaction effect between perceived output quality and provider trustworthiness (not supporting H3), indicating that high output quality did not compensate for the negative impact of low provider trustworthiness on people's disclosure intention. The non-existent interaction is in line with the results of an experimental study by Wottrich et al. (2017), that tested the interaction of perceived benefits (in terms of customization) and brand trust. In contrast to their study, we found a main effect of perceived benefits on disclosure. Thus, our findings extend theirs by suggesting that even when an outcome is perceived as beneficial, it does not necessarily compensate for low provider trustworthiness. Future research should, however, investigate whether system outputs that are perceived as even more beneficial than the ones offered in the current study (i.e., better-personalized recipe recommendations) might still have the potential to rule out negative effects of low provider trustworthiness. Furthermore, it should be studied whether perceptions of provider trustworthiness that are

more related to the treatment of the disclosed information (cf. Bol et al., 2018) than the rather global manipulation of trustworthiness in the current study, could be offset by perceived benefits.

Taken together, our study adds to the utterly limited research investigating (a) the effects of perceived benefits and provider trustworthiness on disclosure intentions within a realistic experimental scenario including actual technology interaction and (b) statistically analyzing the interaction of these factors.

Intention to Use

Adding to the pre-registered analyses, the current study indicated that people's intention to use a recommendation system was related only to perceived output quality but not to provider trustworthiness. Thus, in the context of receiving personalized recommendations, our results propose that when disclosure is less salient (i.e., when asked about the willingness to use rather than the willingness to disclose specific information), people seem to consider provider trustworthiness less, whereas perceived benefits are still relevant. This is, to some extent, in contrast to earlier research showing that personalized services (via perceived benefits) as well as perceived trust (via perceived risks) are related to information disclosure which was in turn closely linked to usage intentions (Pal et al., 2020). One potential explanation for these different findings lays in the context, in which the studies were conducted: while Pal et al.'s study (2020) addressed voice assistants, the technology in the current research was a guide giving recipe recommendations.

Thus, although the intention to use the recipe guide in our study was not attached to the condition of sharing personal information, our research suggests that, in this context, explicit disclosure intentions might depend on different factors than usage decisions which can add to the understanding of people's behavior in the interaction with technologies that collect personal data about their users.

Strengths and Limitations

Our findings are based on a sample of participants that actually interacted with a recommendation system in a realistic scenario before being asked to disclose personal information. Within this interaction, provider trustworthiness was manipulated in a subtle manner using different provider logos and names – with providers being chosen based on the results of a pretest and differences in trustworthiness being supported by a manipulation check. Furthermore, as intended, differences between providers were observed regarding perceived benevolence and integrity, but not regarding competence (see Appendix C for an additional analysis). While we

cannot completely rule out that the chosen type of trustworthiness manipulation (i.e., health insurance vs. food company) might have had unintended side effects, it (1) avoided that participants had to read through study materials that gave explicit information about provider trustworthiness and is (2) more similar to the information people have about the provider in technology interactions in real life. Future studies can strengthen our findings by using other manipulations of provider trustworthiness ensuring not to manipulate any other factors unintendedly. Further, H2, H3, and the respective analysis were pre-registered, and additional exploratory analyses are indicated as such. Thus, this study allows for causal conclusions about the impact of provider trustworthiness on disclosure (as well as usage) intentions. In doing so, the results provide evidence that in disclosure decisions, higher perceived output quality from a recommendation system cannot compensate for a low trustworthiness of its provider, but that provider trustworthiness seems less critical in the decision to use the recommendation system.

Despite the strengths mentioned above, some limitations should be taken into account. In our experimental study, we asked participants about their willingness to disclose specific personal information to a recommendation agent in order to get better-fitting recommendations. It is, however, unclear, in how far our results can be generalized to other contexts that differ, for instance, regarding the amount or intimacy of information that is to be disclosed (and, thus, affect different dimensions of privacy; cf. Burgoon, 1982) or regarding their affordances (e.g., anonymity, persistence, or visibility; cf. Evans et al., 2017), which have been observed to impact disclosure decisions (Trepte et al., 2020). Given that Bol et al. (2019) observed different effects of personalization in different contexts, future studies should address whether our findings replicate in other use contexts. In addition, due to data protection regulations, we could not assess people's actual disclosing behavior. This approach will most likely underestimate the amount of personal data that people would disclose in a real-life interaction as it makes the disclosure of information more salient. Further, we did not experimentally manipulate output quality but only assessed perceived output quality – thus, our results do not allow causal inferences regarding this factor. In addition, the scope of the current research was limited to the interplay between perceived benefits and provider trustworthiness – hence, future studies should include perceived risks (as another important variable in disclosure decisions) and investigate how all three factors play together.

Accordingly, further research should use realistic technology interaction scenarios in different contexts, experimentally manipulate perceived benefits (e.g., output quality), address the interplay of benefits and trustworthiness with perceived risks, and use a different approach

to manipulate provider trustworthiness in order to strengthen the findings from the current study.

Practical Implications

For *technology providers*, the current findings imply that it is not enough to either offer high-quality recommendations or be perceived as trustworthy. Instead, both aspects (i.e., perceived output quality and provider trustworthiness) are related to disclosure, which is an important good for provider companies as it enables insights on their users and offers the possibility for even better-personalized usage experiences and output. Furthermore, as provider trustworthiness seems to be relevant at least for users' conscious disclosure decisions, *regulators* need to ensure that the provider of a technology can be identified without much effort for the user.

Conclusion

Using an experimental approach with actual technology interaction, the current study investigated the interplay of provider trustworthiness and perceived output quality in people's decisions about whether to disclose personal information to a recommendation system. The results indicate that a higher perceived output quality and higher provider trustworthiness are related to a higher willingness to disclose information (if the requested information is relevant to the recommendation). However, both factors seem to act independently, and high perceived output quality does not compensate for low provider trustworthiness. Besides, participants' intention to use the recommendation system depended only on perceived output quality but not on provider trustworthiness. Thus, this study shows that while provider trustworthiness is highly relevant for explicit disclosure decisions and cannot be compensated by high-quality output, it does not impact peoples' decisions to use a recommendation system.

The following chapter contains a manuscript that is the result of a cooperation between Miriam Gieselmann (first author), Dr. Daniel Erdsiek (second author), Vincent Rost (third author), and Prof. Dr. Kai Sassenberg (fourth author). The following table depicts the contributions of the PhD candidate (and of the co-authors, respectively) to the manuscript as well as the current status in the publication process:

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
Miriam Gieselmann	1	50	50	70	60
Daniel Erdsiek	2	10	15	5	10
Vincent Rost	3	10	15	5	10
Kai Sassenberg	4	30	20	20	20
Title of Paper:		Do managers accept Artificial Intelligence? Insights into the role of business sector and AI functionality			
Status in publication process		Submitted for publication			

Chapter 5: Changing the Perspective – AI Acceptance of Decision-Makers

In recent years, artificial intelligence (AI) technologies and applications have become increasingly common in the work context (Kim-Schmid & Raveendhran, 2022). According to a recent survey conducted in the US, the UK, and Germany, about one-third of organizations already use AI applications in several business sectors, and 80% of executives think AI can be applied to any business decision (Rimol, 2022).

AI usage in work(-related) contexts is seen as critical due to multiple reasons (e.g., privacy issues, lack of transparency, or neglect of unique conditions and qualitative information, e.g., Langer, König, & Papathanasiou, 2019; Möhlmann et al., 2021; Newman et al., 2020) but it also offers several benefits. For example, automation is often considered to make more objective decisions than humans (e.g., being more consistent, Langer, König, Sanchez, et al., 2019; less motivated to discriminate, Bigman et al., 2020; and being higher in integrity, Höddinghaus et al., 2021), to provide better decisions (Grove et al., 2000), and to increase efficiency (e.g., in medicine by offering increased diagnostic speed and earlier detection of diseases, Jutzi et al., 2020; or in algorithmic management, Galière, 2020). As these benefits come with the potential to improve organizational attractiveness and effectiveness, as well as to save costs, companies could profit from AI usage.

To unfold these potentials, AI needs to be accepted and implemented in the first place. Thus, it is crucial to understand which factors determine AI acceptance among decision-makers

(i.e., managers). Researchers recently identified that, among others, characteristics of the task (e.g., its objectivity, M. K. Lee, 2018), as well as characteristics of the AI (e.g., its autonomy, Newman et al., 2020), are relevant determinants of acceptance. However, this research predominantly focused on the perspective of people using AI, people being affected by AI, or people observing the consequences of AI usage (for a review, see Langer & Landers, 2021). In contrast, managers' perspective has not been investigated – although they are the ones who ultimately decide whether AI usage should be promoted in a company. To avoid problems arising from potential differences between users' and decision-makers' AI acceptance (e.g., managers deciding to implement AI that is not accepted by their employees or rejecting the implementation of AI that employees would perceive as beneficial), it is necessary to gain a better understanding of managers' perspective.

Thus, the present research aims to add to the literature by investigating determinants of managers' AI acceptance with a focus on two determinants that are likely to influence AI acceptance: (1) the business area in which AI is used – strongly related to task characteristics as will be elaborated below (see, for instance, M. K. Lee, 2018 for an investigation of the impact of task characteristics on users' acceptance) – and (2) AI functionalities (i.e., what the AI is capable of doing) – as a characteristic of the AI (see, for instance, Newman et al., 2020 for a study on the role of AI characteristics on acceptance). Thereby, we sought to provide insights regarding managers' AI acceptance in organizations, which should have some predictive value for future investment decisions made in companies.

AI Acceptance in Work(-Related) Contexts

The limited research on AI acceptance in the work context has identified several factors influencing AI acceptance. A recent review considering the perspectives of (1) people being targeted by automated decisions and (2) observers of such decisions highlights the following determinants for the acceptance of automated/augmented decisions (Langer & Landers, 2021): characteristics of (a) the human evaluating the system (e.g., gender; Dineen et al., 2004), (b) the outcomes (e.g., biased outputs; Hong et al., 2020), (c) the (AI) technology (e.g., transparency; Newman et al., 2020), and (d) the task that is to be completed (e.g., low- vs. high-stake decisions; Langer, König, & Papathanasiou, 2019). Adding to previous research on (c) and (d), other specific characteristics of the AI (e.g., its functionality) and the task (e.g., its objectivity) could be crucial for AI acceptance as well. As will be derived below, it might make a large difference whether AI is used in the *business area* of human resources (HR) or an area in which tasks are typically perceived to be more objective (e.g., finances; cf. Castelo et al., 2019) or in

which personal data is less relevant, such as finances or marketing. In addition, the *functionality* ranging from monitoring a state to autonomously implementing changes might also be relevant.

Role of Business Area

Business areas differ regarding the tasks typically associated with them. Given that AI acceptance depends on task type, this might also have implications for accepting AI in different business areas. People are more willing to accept automation in tasks perceived as relatively objective (e.g., financial advice) than in tasks perceived as more subjective (e.g., dating advice; Castelo et al., 2019). Similar observations have been made in the work context as well. In 'mechanical' tasks (i.e., work assignment, scheduling), humans and computers were perceived as equally fair and trustworthy, whereas in 'human' tasks (i.e., hiring, work evaluation), humans were perceived as fairer and more trustworthy than computers and evoked fewer negative emotions (M. K. Lee, 2018).

An area that is, by definition, associated with 'human' tasks is HR. Here, skills such as empathy and intuition – indicating more subjectivity – are often perceived as useful (e.g., Gikopoulos, 2019), whereas the area of finances requires rule-based, logical, and, thus, more objective decision-making (e.g., stock predictions were perceived as more objective than student performance predictions or hiring decisions; Castelo et al., 2019). This would imply that AI acceptance is lower in HR than in finances.

Furthermore, HR, more than any other business area, deals with personal data of applicants and employees. Given that people are often concerned about privacy when interacting with (AI) technology (e.g., Yan et al., 2022), it seems likely that AI acceptance in HR (compared to other business sectors with lower relevance of personal data, such as finances or marketing) is particularly low.

To our knowledge, no research has explicitly focused on comparing managers' AI acceptance and AI acceptance in general between (these) business areas. Nonetheless, there is initial evidence suggesting that people might be reluctant to use AI in HR: Höddinghaus et al. (2021) report that in disciplinary (i.e., bonus payment) and mentoring (i.e., allocation of training/workshops) tasks, human agents were perceived as more adaptive and benevolent than computers (although computers were perceived as higher in integrity and transparency). In contrast, people seem to be (even overly) willing to accept AI usage in more objective tasks. For investment choices in research and development, people have been observed to over-rely on AI decisions, which might be based on attributing a more structured process for the AI than for a human (Keding & Meissner, 2021). Based on the argumentation above, it seems likely that AI usage in HR is less accepted than in marketing (which typically also includes 'human' tasks but has

a lower relevance of personal data) and finances (typically dealing with less 'human' tasks and having a lower relevance of personal data). Thus, we expected the following:

H1: Usage of AI is less accepted in HR than in other business areas (i.e., finances, marketing).

Role of AI Functionalities

AI can take over different parts of a task. Kaber and Endsley (2004) distinguish four consecutive functionalities (from low to high functionality) of technical systems: *Monitoring* describes the functionality to take all relevant information into account and perceive the current status. *Generation* means the formulation of different options or strategies to achieve goals. *Selection* refers to deciding on one particular option or strategy, while *implementation* means the chosen option is carried out.

At first sight, a higher AI functionality can be regarded as positive. Higher functionality is likely related to higher perceived usefulness, which is considered a crucial determinant of technology acceptance (e.g., Technology Acceptance Model, Davis et al., 1989). Furthermore, empirical studies show functionality is associated with acceptance (e.g., continuance intentions) of AI technologies (e.g., Dehghani, 2018).

However, it seems that technologies can also become too competent (e.g., Gieselmann & Sassenberg, 2023b). Functionality is often no longer seen as positive if the (AI) system acts entirely autonomously and, consequently, human control declines. For example, in medical decision-making, humans often prefer an (AI) system as an informant or second opinion over a system as the ultimate decision-maker (e.g., Jutzi et al., 2020; Palmisciano et al., 2020). More relevant to the current research, also in the work context – for example, for bonus allocation – people perceived augmented decisions (i.e., the human had the option to consult an automated system) as fairer than fully automated decisions and decisions where the human could only slightly change the system output (Newman et al., 2020).

In addition, there is initial evidence that the effect of AI functionality might depend on the business area. In a study that focused on the role of task complexity, it was observed that automated decisions (compared to human and augmented decisions) led to more negative perceptions (of procedural justice) in hiring and performance evaluation (i.e., typical HR tasks). In contrast, in more mathematical tasks (i.e., travel reimbursements or calculating pensions), less human control led to more favorable perceptions (Nagtegaal, 2021). In sum, functionality (like business area) seems to affect AI acceptance. Based on the (limited) existing studies, it is, however, not clear whether functionality leads to more AI acceptance, less AI acceptance, or whether the impact of functionality depends on the business area – functionality might, for

instance, lead to higher acceptance in finances and lower acceptance in HR. Therefore, we did not formulate a hypothesis but a research question regarding the impact of functionality.

RQ: Does functionality influence AI acceptance, and does this effect depend on the business area?

Current Research

We investigated the above-described ideas in a stakeholder group that has not been studied in research so far, namely managers. Different stakeholders have different interests regarding AI usage in the work context (Langer et al., 2021). For managers (i.e., the people who decide whether AI should be used in a company), criteria like efficiency and potential cost savings will most likely be crucial: for example, they might be willing to accept AI usage if it offers the potential to save costs (e.g., by reducing employees' workload; Balfe et al., 2015) while this might cause concerns about job loss among employees. Thus, managers' perspectives on AI usage in companies might essentially differ from those of other groups. However, as outlined above, previous research has mainly focused on the perspective of people using the AI, being affected by AI decisions, or uninvolved observers (for a review, see Langer & Landers, 2021). To address this gap in empirical knowledge, our analyses focus on companies' decision makers (i.e., managers) whose perception and acceptance of AI technologies are crucial for the future diffusion of AI.

In four experimental studies, we varied the business area in which AI is used as well as the functionalities of AI. In the first study, we aimed to grasp an initial understanding of managers' AI acceptance – in terms of perceived risks of and willingness to invest in AI usage in different business areas (namely HR and finances) and for different task functionalities. In Studies 4.2a and 4.2b, we aimed to validate our findings in larger samples with adjusted materials (i.e., using different AI application scenarios). In Study 4.3, we aimed to extend our findings for the comparison of business areas by comparing HR to another business area (i.e., marketing instead of finances). Studies 4.1 and 4.3 were conducted in German-speaking countries, and Studies 4.2a and 4.2b in the UK.

The effect of the business area was pre-registered for all studies; the idea of the RQ was addressed in the preregistrations of all reported studies. All deviations from the preregistrations are reported in detail in Appendix D.

Study 4.1

Method

Design and Participants. We aimed to sample 150 valid cases for this preliminary study (https://aspredicted.org/VQS_2NX). We recruited German-speaking participants with management experience via Prolific and compensated them with 2.50£. Data collection took place in September and October 2022. After exclusion as pre-registered (for details, see Appendix D; this applies for all studies reported in this chapter), the eligible sample consisted of $N = 168$ (gender: 33.92% female, 65.48% male, others preferred not to say; aged: 18-64, $M = 35.61$ years). All participants reported to have management experience. At the time of data collection, 10.7% were part of the top management, 14.3% middle management, 51.8% team/project leaders, 20.8% employees without leadership function, and 2.4% other. The most frequently represented business sectors were information technology (25.6%), education/healthcare/public sector (24.4%), and services (9.5%). Participants worked in the divisions of information technology and electronic data processing (17.3%), research and development (13.7%), top management/executive board (12.5%), manufacturing (10.1%), customer service and sales (each 7.7%), HR (6.0%), marketing (4.8%), finances (3.6%), others (11.3%), 5.4% preferred not to say. The study was implemented as a 2 (business area: HR vs. finances, between-subjects) x 4 (functionality: monitoring, generation, selection, implementation, within-subjects) mixed design.

Procedure. Participants were randomly assigned to one of two conditions representing AI usage in different business areas: HR or finances. After a brief introduction, participants read a short case study of a company aiming to use AI in the respective business area. Within the case study, four consecutive tasks with increasing functionality of the AI (i.e., monitoring, generation, selection, implementation) were presented to participants (see Table 10). After each task description, participants were asked about (a) the perceived risk and (b) their willingness to invest in AI applications for the described task. A complete list of all measures taken, instructions, and items is provided in Appendix D – this applies to all studies reported in this chapter.

Table 10

Described AI Functionalities in HR and Finances (Study 4.1)

Functionality	HR	Finances
Monitoring	Analyze data from previous application processes and identify characteristics of successful employees	Analyze past performance data of marketing channels and identify efficient investment strategies
Generation	Propose recruitment channels fitting the characteristics of successful employees	Propose different investment strategies
Selection	Select the most appropriate (combination of) recruitment channel(s)	Select the most appropriate investment strategies
Implementation	Generate and publish content on selected recruitment channels	Inform responsible persons about the available budget for their channel

Note. These are shortened and translated versions of the material. The original material is provided in Appendix D.

Measures. We measured *perceived risk* ('Using AI for this task comes with the risk of negative consequences for my company.', $M = 3.99$, $SD = 1.10$) and *willingness to invest* ('My company should invest to increase usage of AI for this task.', $M = 4.53$, $SD = 1.26$) with one item each assessed on seven-point Likert scales (1 = *strongly disagree*, 7 = *strongly agree*). Both measures were taken separately for each level of AI functionality.

Results

We calculated mixed ANOVAs with business area (finances vs. HR) as a between-factor, functionality as a within-factor, and (a) perceived risk of negative consequences and (b) willingness to invest as dependent variables. In case of a main effect of functionality (monitoring, generation, selection, implementation), we used contrast-coding in the follow-up analysis (C1: -1/4, -1/4, -1/4, 3/4; C2: -2/3, 1/3, 1/3, 0; C3: 0, -1/2, 1/2, 0).

Risk of Negative Consequences. An HF-corrected ANOVA with perceived risk as dependent variable (for descriptive results, see Table 11) showed no difference between HR and finances, $F(1, 166) = 1.06$, $p = .304$, $\eta_p^2 = 0.01$, but revealed a significant interaction between functionality and business area, $F(2.84, 471.79) = 5.55$, $p = .001$, $\eta_p^2 = 0.03$. Thus, we calculated separate ANOVAs for both business areas.

Table 11

Willingness to Invest and Risk of Negative Consequences for Different AI Functionalities in HR and Finances in Study 4.1 (N = 168)

Functionality	Monitoring		Generation		Selection		Implementation	
Area	HR	Finances	HR	Finances	HR	Finances	HR	Finances
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Risk	3.89 (1.35)	3.35 (1.43)	3.62 (1.44)	3.83 (1.51)	3.94 (1.45)	4.10 (1.58)	4.86 (1.51)	4.33 (1.59)
Investment	4.13 (1.67)	4.83 (1.35)	4.57 (1.59)	4.81 (1.18)	4.40 (1.52)	5.01 (1.27)	4.04 (1.66)	4.49 (1.44)

Note. For brevity, we refer to the risk of negative consequences as *risk* and willingness to invest as *investment* in this table.

For HR, perceived risk was lower for monitoring, generation, and selection compared to implementation (C1), $F(1, 83) = 39.19$, $p < .001$, $\eta_p^2 = 0.32$. Perceived risk did neither differ for monitoring compared to generation and selection (C2), $F(1, 83) = 0.57$, $p = .453$, $\eta_p^2 = 0.01$, nor for generation compared to selection (C3), $F(1, 83) = 3.79$, $p = .055$, $\eta_p^2 = 0.04$.

For finances, perceived risk was also lower for monitoring, generation, and selection compared to implementation (C1), $F(1, 83) = 14.79$, $p < .001$, $\eta_p^2 = 0.15$. Besides, perceived risk was lower for monitoring compared to generation and selection (C2), $F(1, 83) = 17.84$, $p < .001$, $\eta_p^2 = 0.18$, while there was no difference between generation and selection, $F(1, 83) = 2.57$, $p = .113$, $\eta_p^2 = 0.03$.

Willingness to Invest. The HF-corrected ANOVA with willingness to invest as dependent variable (for descriptive results, see Table 11) revealed no significant interaction between functionality and business area, $F(2.54, 408.21) = 2.11$, $p = .109$, $\eta_p^2 = 0.01$. Willingness to invest was lower in HR than in finances, $F(1, 166) = 6.81$, $p = .010$, $\eta_p^2 = 0.04$. In addition, willingness to invest differed depending on functionality, $F(2.54, 408.21) = 8.99$, $p < .001$, $\eta_p^2 = 0.05$.

Across business areas, willingness to invest was higher for monitoring ($M = 4.48$, $SD = 1.56$), generation ($M = 4.69$, $SD = 1.40$), and selection ($M = 4.71$, $SD = 1.43$) compared to implementation ($M = 4.26$, $SD = 1.56$) (C1), $F(1, 167) = 13.44$, $p < .001$, $\eta_p^2 = 0.07$, but lower for monitoring compared to generation and selection (C2), $F(1, 167) = 8.39$, $p = .004$, $\eta_p^2 = 0.05$.

Willingness to invest did not differ between generation and selection (C3), $F(1, 167) = 0.04$, $p = .832$, $\eta_p^2 < 0.01$.

Discussion

The results of Study 4.1 show that managers' willingness to invest is lower for AI applications in the HR area than in the finances area. At the same time, perceived risk did not differ significantly between these business areas. Both aspects were further dependent on AI functionality. A general trend was observable that higher functionality – especially the highest functionality *implementation* – leads to less acceptance (i.e., higher perceived risk and less willingness to invest). However, the HR monitoring condition (i.e., the lowest functionality) differed from this pattern (i.e., relatively high perceived risk and low willingness to invest). This deviation might be due to a confound in the study material: the described AI usage in monitoring (but not the other levels of functionality) included the analysis of application data (i.e., highly sensitive personal data), and it was unclear whether this data would be anonymized. Another limitation of Study 4.1 is the small sample size. We, thus, aimed to validate our findings in a larger sample with adjusted materials (including a different description of AI functionalities for both business areas) in Studies 4.2a and 4.2b.

Studies 4.2a and 4.2b

Method

Design, Participants and Procedure. Studies 4.2a and 4.2b were conceptual replications of Study 4.1 with adjusted materials. The main difference between Studies 4.2a and 4.2b was a change in materials for HR implementation (see Table 12). For both studies, we recruited native English speakers living in the UK, having management experience, and working at least 20 hours per week via Prolific.

Participants in Study 4.2a. We aimed to sample 400 valid cases for this study (https://aspredicted.org/VF6_L74) to reach a power of 90% and an alpha error of 0.05 for observing an effect of $f \geq 0.15$ for the between-subjects factor in a mixed ANOVA with four measures and two groups. Data collection took place in December 2022. Participants were compensated with 2.50£ for their participation. After exclusion as pre-registered, the eligible sample consisted of $N = 451$ valid cases (gender: 39.02% female, 54.77% male, 0.67% non-binary, others preferred not to indicate; aged: 20-70, $M = 40.88$ years). All participants reported to have management experience. At the time of data collection, 8.0% were part of the top management, 29.5% middle management, 43.2% team/project leaders, 17.5% employees without leadership

function, and 1.8% other. The most frequently represented business sectors were education/healthcare/public sector (27.5%), finances/consulting/insurances (12.0%), information technology (11.5%), manufacturing (9.1%), and services (7.5%). Participants worked in the divisions of information technology (15.5%), customer service (14.0%), research and development (11.1%), finances (10.6%), top management/executive board (12.5%), sales (7.1%), HR (4.2%), manufacturing (4.2%), marketing (3.6%), others (20.0%).

Participants in Study 4.2b. Based on a sequential testing approach, we pre-registered (https://aspredicted.org/CSK_3J7) to initially collect data from 200 participants. If this first data collection revealed a non-significant effect of $f \geq 0.15$ for the main effect of business area or the interaction between the business area and functionality in the pre-registered mixed ANOVAs, we would have collected additional data. As the described effects were significant based on the first data collection, we did not collect any additional data. Data collection took place in January 2023. Participants were compensated with 2.10£ for their participation. After exclusion as pre-registered, the eligible sample consisted of $N = 186$ valid cases (gender: 40.86% female, 57.53% male, 0.54% non-binary, others preferred not to indicate; aged: 20-71, $M = 42.35$ years). All participants reported to have management experience. At the time of data collection, 5.9% were part of the top management, 31.7% middle management, 37.6% team/project leaders, 22.0% employees without leadership function, and 2.7% other. The most frequently represented business sectors were education/healthcare/public sector (31.2%), information technology (16.1%), finances/consulting/insurances (12.4%), and retail (8.1%). Participants worked in the divisions of customer service (16.1%), information technology (11.8%), top management/executive board (9.7%), finances/research and development (8.1% each), HR (7.0%), sales (5.4%), marketing (4.8%), manufacturing (1.1%), others (28.0%).

Measures. In both studies, we assessed perceived risk ('Using AI for this task comes with the risk of negative consequences for my company,' $M_{\text{Study 4.2a}} = 35.50$, $SD_{\text{Study 4.2a}} = 27.95$, $M_{\text{Study 4.2b}} = 41.66$, $SD_{\text{Study 4.2b}} = 20.88$) and willingness to invest ('My company should invest to increase the usage of AI in this task,' $M_{\text{Study 4.2a}} = 39.90$, $SD_{\text{Study 4.2a}} = 30.05$, $M_{\text{Study 4.2b}} = 42.64$, $SD_{\text{Study 4.2b}} = 26.31$) with one item each using sliders from 0 to 100 (0 = *very low risk of negative consequences/ not invest at all*, 100 = *very high risk of negative consequences/ invest as much as possible*). These measures were taken for each of the four levels of functionality.

Table 12

Described AI Functionalities in HR and Finances (Study 4.2a and 4.2b)

Functionality	HR	Finances
Monitoring	Evaluate anonymized application/ performance data and highlight characteristics of successful employees	Evaluate company's past cash flow data and highlight indicators of suspicious transactions
Generation	Scan work-related social media profiles and rate their suitability for vacancy	Scan current cash flow and rate suspiciousness of each transaction
Selection	Select persons to be considered in the further recruitment process	Flag highly suspicious transactions
Implementation	*Study 2a:* Invite selected persons to apply by sending messages that they have been identified as suitable candidates *Study 2b:* Decide whom to invite to an online assessment that serves as first step in the personnel selection process	Decide about (non-)execution of transactions: Put suspicious transactions on hold and request permission from supervisor

Note. These are shortened versions of the material. The original material is provided in Appendix D.

Results

We calculated the same mixed ANOVAs as in Study 4.1. Perceived risk of negative consequences (see Table 13 for descriptive results) was higher in HR than in finances in Study 4.2a, $F(1, 449) = 96.46$, $p < .001$, $\eta_p^2 = 0.18$, and Study 4.2b, $F(1, 184) = 35.29$, $p < .001$, $\eta_p^2 = 0.16$. Willingness to invest (see Table 13 for descriptive results) was lower in HR than in finances in Study 4.2a, $F(1,449) = 27.64$, $p < .001$, $\eta_p^2 = 0.06$, and Study 4.2b, $F(1, 184) = 37.63$, $p < .001$, $\eta_p^2 = 0.17$. Due to significant interactions between functionality and business area for both dependent variables (all ps $< .02$; all η_p^2s > 0.01), we calculated follow-up analyses separately for finances and HR to investigate the effects of functionality (see Table 14 and below for results).

Table 13

Willingness to Invest and Risk of Negative Consequences for Different AI Functionalities in HR and Finances in Studies 4.2a (N = 451) and 4.2b (N = 186)

Functionality	Monitoring		Generation		Selection		Implementation	
Area	HR	Finances	HR	Finances	HR	Finances	HR	Finances
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Study 4.2a								
Risk	46.07 (27.86)	26.24 (21.90)	48.38 (29.28)	24.10 (21.95)	46.20 (28.57)	22.13 (20.89)	39.03 (28.20)	31.66 (28.47)
Investment	36.25 (26.93)	42.94 (28.66)	29.54 (27.44)	42.74 (29.80)	31.31 (26.99)	52.52 (31.55)	37.55 (29.78)	46.45 (32.20)
Study 4.2b								
Risk	48.02 (27.15)	31.62 (23.89)	55.84 (29.89)	32.72 (25.26)	51.43 (29.60)	30.74 (27.17)	36.08 (25.24)	55.95 (32.11)
Investment	34.10 (24.76)	49.84 (29.41)	25.30 (24.10)	49.82 (30.99)	31.84 (26.65)	58.22 (32.42)	44.75 (27.99)	38.13 (28.41)

Note. For brevity, we refer to the risk of negative consequences as *risk* and willingness to invest as *investment* in this table.

Table 14

Results of Mixed ANOVAs for Risk of Negative Consequences and Willingness to Invest in HR and Finances in Study 4.2a (N = 451) and Study 4.2b (N = 186)

		Risk of negative consequences				Willingness to invest			
Predictor	Study	*F*	*df*	*p*	η_p^2	*F*	*df*	*p*	η_p^2
Follow-up analysis: HR									
C1 (MGS vs. I) [a]	4.2a	**18.93**	**1, 225**	**< .001**	**0.08**	**11.03**	**1, 225**	**.001**	**0.05**
	4.2b	**6.09**	**1, 92**	**.015**	**0.06**	**5.00**	**1, 92**	**.028**	**0.05**
C2 (M vs. GS) [b]	4.2a	0.64	1, 225	.425	< 0.01	**16.96**	**1, 225**	**< .001**	**0.07**
	4.2b	2.56	1, 92	.113	0.03	**5.34**	**1, 92**	**.023**	**0.05**
C3 (G vs. S) [c]	4.2a	1.47	1, 225	.226	0.01	1.63	1, 225	.203	0.01
	4.2b	2.01	1, 92	.160	0.02	**10.07**	**1, 92**	**.002**	**0.10**
Follow-up analysis: Finances									
C1 (MGS vs. I) [a]	4.2a	**19.29**	**1, 224**	**< .001**	**0.08**	0.06	1, 224	.801	< 0.01
	4.2b	**4.78**	**1, 92**	**.031**	**0.05**	2.38	1, 92	.126	0.03
C2 (M vs. GS) [b]	4.2a	**8.17**	**1, 224**	**.005**	**0.04**	**15.79**	**1, 224**	**< .001**	**0.07**
	4.2b	0.00	1, 92	.962	< 0.01	**4.20**	**1, 92**	**.043**	**0.04**
C3 (G vs. S) [c]	4.2a	2.41	1, 224	.122	0.01	**54.43**	**1, 224**	**< .001**	**0.20**
	4.2b	0.44	1, 92	.508	0.01	**13.42**	**1, 92**	**< .001**	**0.13**

[a] Monitoring (M)/generation (G)/selection (S) = -1/4, implementation (I) = 3/4.

[b] Monitoring = -2/3, generation/selection = 1/3, implementation = 0.

[c] Monitoring = 0, generation = -1/2, selection = 1/2, implementation = 0.

Risk of Negative Consequences. For HR, perceived risk was higher for monitoring, generation, and selection, compared to implementation (C1), while the other two contrasts (C2 and C3) indicated no significant differences. For finances, perceived risk was lower for monitoring, generation, and selection compared to implementation (C1), but there was no difference between generation and selection (C3). In Study 4.2a, perceived risk was higher for monitoring than generation and selection (C2), while the perceived risk of these functionalities did not differ in Study 4.2b.

Willingness to Invest. For HR, willingness to invest was lower for monitoring, generation, and selection compared to implementation (C1) and higher for monitoring compared to generation and selection (C2). Furthermore, in Study 4.2b, willingness to invest was lower for generation than for selection (C3), while there was no significant difference between these two functionalities in Study 4.2a. For finances, willingness to invest was lower for monitoring

compared to generation and selection (C2) and lower for generation compared to selection (C3). There was no significant difference between monitoring, generation, and selection compared to implementation (C1).

Discussion

Results of Study 4.2a and 4.2b show that managers perceive a higher risk of AI usage in HR compared to finances (different from Study 4.1, where no difference was observed) and are less willing to invest in AI applications in HR (in line with Study 4.1). This indicates that the HR area might be especially critical. To provide additional evidence that HR (and the high relevance of personal data), rather than finances (with more objective tasks), drives this effect, we replaced finances with another business area (i.e., marketing, which typically has a lower relevance of personal data than HR but typical tasks can be considered to require inherently human skills) in Study 4.3.

As in Study 4.1, perceived risk and willingness to invest differed depending on AI functionality in Studies 4.2a and 4.2b, but results here were more mixed and in part inconsistent with Study 4.1. The idea that higher functionality leads to less acceptance was only partially supported: in HR, participants were more willing to invest in monitoring than in generation and selection, and in finances, perceived risk was highest for implementation. However, for HR implementation, perceived risk was surprisingly low, and willingness to invest was relatively high. To rule out that this effect was due to the relatively soft description of HR implementation in Studies 4.2a and 4.2b (invite potential candidates to apply for vacancies or to a short online assessment), we used a stronger manipulation of AI implementation in HR in Study 4.3, indicating clearly that no human employee is involved in the implementation or controls the AI in advance of implementation.

Besides, in finances, willingness to invest was relatively high for AI selection, indicating that managers might see more benefits in higher AI functionalities and, thus, be more willing to invest in such applications – as long as the AI functionalities do not cross a critical level (as is the case for implementation). To ensure that the previously described findings for functionality are not an artifact of the experimental design (functionality as a non-randomized within-subjects factor), we chose a different design in Study 4.3 (functionality as a between-subjects factor). Furthermore, we aimed to validate our findings with a larger sample of managers and thus distributed our survey to a large number of German companies in Study 4.3. As this recruitment channel required a brief survey, we focused on two of the four functionalities: generation and implementation. This selection was guided by the fact that the implementation

condition differed from the other conditions in most of the analyses, and generation (other than monitoring) seemed to be consistently contributing to this difference.

Study 4.3

Method

Design and Participants. Study 4.3 (https://aspredicted.org/8SL_WYL) was implemented as a 2 (business area: HR vs. marketing, within-subject) x 2 (functionality: generation vs. implementation, between-subjects) mixed-measures design. Participants were recruited via a quarterly survey regularly distributed to German companies (at least five employees; knowledge-intensive service area and manufacturing industry). Data collection took place in April and May 2023. After exclusion as pre-registered, the eligible sample consisted of N = 1,220 (aged: 1.72% under 30, 7.87% 30-39, 17.54% 40-49, 44.26% 50-60, 27.38% >60 years, others did not indicate their age). At the time of data collection, 74.18% of participants were CEOs/part of the executive board, 16.56% were department heads (or higher management), 2.70% were team/project leaders, and 7.13% were employees without management function. Participants worked in the divisions of top management/executive board (77.0%), HR (14.2%), marketing (8.3%), information technology (5.1%), others (18.4%).

Procedure. Companies were randomly assigned to receive a survey version describing application scenarios of either (a) generating AI or (b) implementing AI for two business areas (HR and marketing, see Table 15). Afterward, participants were asked whether they would use such an AI tool in their company.

Table 15

Described AI Functionalities in HR and Marketing (Study 4.3)

Functionality	HR	Marketing
Generation	Scan work-related social media profiles and rate their suitability for the vacancy, HR team has access to all ratings and can choose whom to invite for an interview	Analyze data from previous marketing campaigns and generate new marketing content, marketing team has access to variety of AI-generated content and decides about publication
Implementation	Scan work-related social media profiles and rate their suitability for the vacancy, invite suitable candidates for an interview without a review by the HR team	Analyze data from previous marketing campaigns and generate new marketing content, AI-generated content is autonomously published without review by the marketing team

Note. These are shortened versions of the material. The original material is provided in Appendix D.

Measures. *Willingness to use* (M = 4.63, SD = 2.61) was assessed with one item ('Would you use such an AI tool in your company?', 10-pt scale from 1 = 'no, by no means' to 10 'yes, definitely'). This measure was separately collected for both business areas. We considered willingness to use a more direct measure of acceptance than willingness to invest because the latter (but not the former) depends on financial resources.

Results

We calculated a mixed ANOVA with business area (HR vs. marketing) as a within-factor, functionality (generation vs. implementation) as a between-factor, and willingness to use as the dependent variable, see Table 17. Willingness to use AI was lower for HR than for marketing (for descriptive results, see Table 16). In addition, it was higher for generation than for implementation. There was no significant interaction between the business area and the functionality.

Table 16

Willingness to Use Different AI Functionalities in HR and Marketing in Study 4.3 (N = 1220)

Functionality	Generation		Implementation	
Area	HR	Marketing	HR	Marketing
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Willingness to use	4.78 (2.96)	5.12 (3.03)	4.20 (2.83)	4.43 (3.00)

Table 17

Results of Mixed ANOVA for Willingness to Use in HR and Marketing in Study 4.3 (N = 1220)

	Willingness to use			
Predictor	*F*	*df*	*p*	η_p^2
Area	**12.25**	**1, 1218**	**< .001**	**0.01**
Functionality (G vs. I)	**18.30**	**1, 1218**	**< .001**	**0.02**
Area x Functionality	0.44	1, 1218	.508	< 0.01

Discussion

Study 4.3 shows that managers are less willing to use AI in HR than in marketing – accordingly, the findings from Studies 4.1, 4.2a, and 4.2b seem not to be limited to comparing HR and finances. Instead, AI usage in HR might be seen as especially critical (also in comparison to AI usage in marketing). Accordingly, the relevance of personal data and associated concerns might constitute another explanation for low AI acceptance in HR that goes beyond the requirement of human skills (that applies to both HR and marketing). Furthermore, in both business areas, willingness to use was higher for generation than implementation. This supports the idea that highly functional AI (i.e., AI that implements) is less accepted than AI with lower functionalities, replicating the findings of Study 4.1.

Discussion of Chapter 5

The current research tested whether managers' AI acceptance (in terms of perceived risks, willingness to invest, and willingness to use) depends on the business context in which AI is used (namely HR vs. finances/ marketing) and, moreover, considered the role of AI functionality. We predicted managers' acceptance of AI applications would be lower in HR than in finances/ marketing (H1). We further explored the role of AI functionalities and the interplay of business area and AI functionality (RQ).

Role of Business Area

Across four studies, we found evidence that managers are less willing to accept AI usage in HR than in finances and marketing. This finding adds to the literature investigating AI acceptance in specific business contexts such as HR (Höddinghaus et al., 2021) and investment choices (Keding & Meissner, 2021). Despite the consistent results within our research, further studies are needed replicating the observed differences between business areas with other materials (e.g., vignettes describing AI that completes other tasks in the respective business areas), to rule out that the observed effects are limited to the tasks described within our studies.

The observed differences between HR and finances are in line with earlier research showing lower acceptance of automation in more subjective and 'human' tasks than in more objective and mechanical tasks (Castelo et al., 2019; M. K. Lee, 2018). Typical HR tasks (e.g., recruiting) might be considered as tasks that humans can complete better than machines – for example, because they are perceived to be more subjective (i.e., not objectifiable and rule-based) and to require more skills that are inherently human but lacking in machines (e.g., intuition and empathy) than typical tasks in the finances sector (e.g., executing transactions). However, we also observed differences between HR and marketing – a sector where typical tasks (e.g., the creation of a marketing campaign) might also be perceived as rather subjective and as requiring inherently human skills. Thus, another potential explanation might be more useful to explain the observed effects: in every company, HR deals – more than any other business area – with highly sensitive personal data of employees and applicants. Thus, particularly in HR, privacy concerns might hamper acceptance. Accordingly, the relevance of personal data in certain business sectors might be crucial for managers' AI acceptance. At this point, these explanations are both speculative and should, thus, be tested in future research.

Role of AI Functionality

Findings for AI functionality were less conclusive. For the lower functionalities monitoring, generation, and selection, we observed different acceptance patterns across studies. Studies 4.2a and 4.2b hint at the possibility that in finances, managers might be more willing to accept AI the more functional it is – as long as functionality stays below a critical point (i.e., implementation). However, this pattern is not observable in Study 4.1 and not in the business area of HR. Thus, this observation can only be seen as an initial hint and should be interpreted cautiously. Accordingly, we highly encourage researchers to investigate the effects of AI functionality in different business areas more thoroughly.

Despite the inconsistencies in comparing lower functionalities, it becomes evident that acceptance of AI usage in both areas declines as soon as AI implements without allowing a

human to intervene in advance. The comparably high acceptance of HR implementation in Studies 4.2a and 4.2b can be explained by the chosen materials, which described implementation (i.e., inviting people to apply or for a short online assessment) without making explicit that the AI acts completely autonomously without offering the human to intervene. Thus, it seems that consequential implementation without a-priori human validation is detrimental to AI acceptance irrespective of the business area. Nonetheless, further studies are needed to systematically investigate the importance of human intervention options for the acceptance of AI that implements autonomously – for instance, by manipulating this factor experimentally.

Strengths and Limitations

The reported studies are the first to experimentally investigate managers' perspectives on AI usage in their companies. Our findings allow first insights into the role of the business area and AI functionality on AI acceptance of this important group, which had – up to now – not received the necessary attention in the literature. Based on data from several experimental studies with a large overall sample size, the results suggest that managers are less willing to accept AI usage in HR than in other business areas (i.e., finances/ marketing). Furthermore, the results allow first cautious conclusions about the role of AI functionality – namely, a presumable decline in AI acceptance as soon as functionality crosses a critical boundary.

Despite the strengths of the current research, some limitations should be considered when interpreting the results. Our experimental studies were based on very brief and generic hypothetical scenario descriptions. Participants received neither very detailed information about the AI system in general nor about how it could be applied in their own company. We opted for the brief materials given that we were studying managers with a tight schedule. If feasible, it would be beneficial if future research used more detailed descriptions of AI applications in companies. Furthermore, although we tried to keep the functionality descriptions between the investigated business areas as equivalent as possible, we cannot entirely rule out that they were perceived differently by participants regarding certain aspects (e.g., the perceived autonomy of the AI). Besides, we relied on self-reported acceptance measures (i.e., perceived risk of negative consequences and willingness to invest). Although intentions are generally a good predictor of behavior, future research should also investigate actual company investments in AI and usage of AI.

Conclusion

The present research investigated the role of the business area and AI functionality on managers' AI acceptance in companies. While managers' perspective had not been investigated

until now, our findings propose that they are less willing to accept AI usage in HR than in other areas (i.e., finances/ marketing). Moreover, it appears that managers are – irrespective of the business area – (currently) unwilling to accept very high functions of AI (i.e., implementation) that include potential consequences that cannot be prevented upfront by a human.

Chapter 6: General Discussion

The present dissertation enhances our understanding of disclosure in increasingly common human-AI interactions by investigating it from a human-oriented perspective. In doing so, the thesis focuses on humans' perception of the technology along human-like characteristics rather than the technical features of the technology – based on the ideas of the CASA paradigm (Nass & Moon, 2000; Reeves & Nass, 1996) and anthropomorphism (Epley et al., 2007). More specifically, this thesis examined perceptions of (1) the *interaction partner* and (2) the *interaction* as determinants of disclosure towards (and acceptance of) AI technologies, as well as the relevance of information disclosure for AI acceptance.

In the first three empirical chapters (i.e., Chapters 2 to 4), the dissertation focused on determinants of users' disclosure in private use contexts of AI. More specifically, it addressed the role of (1) specific perceptions of the technology and the provider as *interaction partners* and (2) perceived interactivity and output quality as characteristics of the *interaction*.

Chapter 2 investigated the relationship between task-related characteristics of the *technology* as interaction partner and disclosure. In detail, this chapter focused on how perceptions of different competence levels (derived from the action regulation theory; Hacker, 1971, 1998) in conversational AI affect privacy concerns and the general willingness to disclose personal data. Across two survey studies and three vignette experiments, the results suggest that perceived competencies of conversational AI have differential effects for non-users and users of the technology. Non-users were indifferent to intellectual competencies of conversational AI, but higher perceived competencies in terms of meta-cognitive heuristics raised their privacy concerns and decreased their willingness to share personal data with the system. In contrast, for users, intellectual competencies were associated with a higher willingness to disclose and lower privacy concerns. Furthermore, for users, higher meta-cognitive heuristics were related to more privacy concerns but not to a lower willingness to disclose. Taken together, the findings of Chapter 2 highlight the importance of task-related perceptions of the technology (i.e., its perceived competencies) as determinants of disclosure in human-AI interactions and stress that more task-related capabilities of AI do not always lead to more positive reactions.

Chapter 3 investigated how perceiving the usage of conversational AI as *interactive* is related to disclosure (i.e., privacy concerns and general willingness to disclose, as in Chapter 2). Differentiating between active control, reciprocal interaction, and synchronicity, two survey studies consistently showed that all these facets of perceived interactivity were positively related to disclosure. However, reciprocal interaction seems more important than the other two facets. While future research can strengthen these findings through experimental support (and

thus allowing causal conclusions), the consistent results across studies of Chapter 3 stress the potential relevance of perceiving interactivity – and, in particular, a reciprocal interaction – for disclosure in human-AI interactions.

Chapter 4 studied the impact of socio-emotional aspects of *provider trustworthiness* and *perceived output quality* on disclosure. In order to do so, the study used an experimental design that included an actual interaction with an algorithm-based recipe guide. The results of Chapter 4 indicate that a less (vs. more) trustworthy provider reduced participants' willingness to disclose specific personal information in order to get better personalized recipe recommendations after an initial interaction with the system. Furthermore, perceiving the quality of the received output (i.e., the recipe recommendations) as higher was associated with a higher willingness to disclose personal information afterward. There was no statistically significant interaction between provider trustworthiness and perceived output quality, demonstrating the independent relevance of both factors for disclosure. Given these results, Chapter 4 sheds light on the distinct importance of socio-emotional perceptions of the interaction partner (i.e., provider trustworthiness with a focus on benevolence and integrity) and perceptions of the interaction (i.e., perceived output quality) for disclosure in human-AI interactions.

In *Chapter 5*, the study focus shifted from individuals using AI systems in private use contexts (as in the previous empirical chapters) to persons deciding about AI usage in contexts in which it affects large groups of people. More specifically, the four experimental studies of Chapter 5 considered the perspective of managers as decision-makers in the work context by studying the role of *AI functionality*. Furthermore, the studies investigated the relevance of personal *data disclosure* by addressing different business areas, which are more (in the case of human resources) or less (in the case of finances and marketing) related to personal data. Accordingly, this chapter considered (the need for) data disclosure as a determinant of acceptance and, thus, extends the findings from the previous empirical chapters focusing on disclosure as an outcome. Results regarding the functionality of AI were somewhat ambiguous across studies for lower functionalities. However, they indicated that managers are currently unwilling to accept AI systems that implement consequential actions (without a-priori human validation). Moreover, managers were more skeptical (i.e., lower willingness to invest and use) towards AI in human resources (i.e., a business area strongly related to personal data) than in other business areas (i.e., finances and marketing). Taken together, Chapter 5 highlights the potential relevance of the necessity to disclose personal data for decision-makers' AI acceptance in work-related contexts. Furthermore, mirroring the results of Chapter 2, it points to the importance of characteristics of the AI as interaction partner (i.e., its functionalities) for disclosure. More

specifically, it became evident that – also in the business context – higher task-related capabilities of AI do not necessarily evoke more positive reactions (as observed for private use contexts of AI in Chapter 2).

In sum, the present dissertation aimed at a better understanding of disclosure in human-AI interactions by using a versatile approach addressing (a) determinants as well as consequences of disclosure, (b) general willingness to disclose and willingness to disclose specific information, (c) different AI technologies, (d) different methodological approaches, and (e) different use contexts of AI. Within a large and heterogenous set of studies, the human-oriented approach proved useful by showing the role of perceptions of (1) the *interaction partner* (i.e., the technology and the provider) and (2) the *interaction* (i.e., perceived interactivity and output quality) as determinants of individuals' *disclosure* in private use contexts of AI. Furthermore, the present thesis highlights the potential relevance of required information disclosure and characteristics of the AI for managers' *AI acceptance*. Taken together, the empirical findings of the current dissertation point out that higher capabilities of an AI can evoke positive reactions, but if they cross a critical boundary, adverse reactions might occur (i.e., negative effects on disclosure and acceptance observed in Chapters 2 and 5). Thus, the results imply that higher AI capabilities might not always be perceived as desirable. This implication will be elaborated more thoroughly in this general discussion after (1) reviewing the general strengths and limitations of the current dissertation and (2) discussing the human-oriented approach to investigating human-AI interactions.

Strengths and Limitations

Strengths

A large number of previous studies have investigated people's privacy concerns and willingness to disclose personal data, focusing in particular on the privacy paradox phenomenon (for a review, see Kokolakis, 2017) and its potential explanations, including the idea of a privacy calculus (e.g., Dinev & Hart, 2006). This dissertation adds to the understanding of disclosure in several aspects. First, studies on this topic have mainly considered the contexts of e-commerce and social media (cf. Kokolakis, 2017). Given that recent research has pointed to differences in disclosure depending on the context (Bol et al., 2018), the current dissertation adds to the existing literature by addressing disclosure not only in private use of different technologies (i.e., conversational AI, recipe recommendation system) but also in work-related contexts. Second, previous studies have predominantly focused on settings in which technology typically served as a tool (e.g., to enable online shopping) or as a communication medium (i.e.,

to enable communication between human interaction partners). Modern technologies, however, are perceived to become more and more agentic (e.g., Glikson & Woolley, 2020). Thus, technology is no longer a mere tool or communication medium but can be regarded as a distinct interaction partner – whose characteristics might impact disclosure (cf. Guzman, 2019). By investigating perceptions of the AI as interaction partner, the current dissertation addresses this increased agency of technology. Third, by considering specific perceptions of the interaction partner and the interaction, this dissertation enables a more fine-grained understanding of which factors influence disclosure in human-AI interactions – adding to the rather broad categories of perceived benefits and costs as addressed in the privacy calculus (e.g., Dinev & Hart, 2006).

Furthermore, this dissertation enhances previous research on disclosure regarding several methodological aspects. Most research on disclosure is based on survey studies – often limited to convenience samples of students – and experiments that lack a realistic context (cf. Kokolakis, 2017), which challenges the external validity of results. All chapters of the current dissertation included more diverse samples than typical student samples and were recruited from relevant target groups (e.g., users of conversational AI or managers as decision-makers in the work context). Besides, Chapter 4 constitutes an example of how to conduct research on disclosure in a realistic setting (i.e., within the usage of a recommendation system close to everyday life) – thus also overcoming typical shortcomings of studies in the field of human-AI interaction (cf. Greussing et al., 2022). Despite the efforts associated with implementing technologies such as the used recipe guide, studies including the interaction with experimentally manipulable technologies offer important insights into human-AI interactions by allowing for causal conclusions from realistic experimental settings and should, thus, be promoted and valued in future research.

Limitations

The research conducted within this thesis has several general limitations that need to be considered when interpreting the results. First of all, the reported studies relied on self-reported measures of disclosure intentions and did not assess actual disclosing behavior. While the decision to assess disclosure intention rather than actual disclosing behavior was made in order to protect participants' privacy (i.e., by avoiding the collection of personal data), this approach poses a limitation to the reported findings due to two reasons: (1) disclosure intentions differ from actual disclosing behavior and (2) explicitly asking participants about their disclosure intentions makes the disclosure more salient than in typical human-technology interactions. First, disclosure intentions are typically lower than actual disclosure (e.g., Norberg et al., 2007). This observation is alarming as it implies that people often disclose more information

than they actually want or intent to. As it is – from a normative perspective – desirable that people only disclose information if they actually want to do so, it is not enough to consider actual disclosing behavior. Instead, to get a comprehensive understanding of disclosure decisions, we also need to understand which factors determine people's disclosing intention (addressed in the current dissertation). Secondly, while information is sometimes explicitly requested in real-life human-technology interactions (making the disclosure similarly salient as in the conducted studies), most personal data is collected implicitly (e.g., by monitoring transactions or usage behavior). Therefore, future research would benefit from studies assessing *actual* disclosure less explicitly than in the current work – and thus closer to realistic human-technology interactions – to ensure that the factors identified as relevant for disclosure intentions in the current thesis are also crucial for actual disclosing behavior.

Secondly, as discussed in the respective chapters, the results regarding users in Chapter 2 and the results of Chapter 3 do not allow for causal conclusions as the reported associations are based on survey data and, thus, correlational. Based on this approach, it cannot be ruled out that the observed relationships are influenced by third variables not included in the current investigations, such as certain motives or characteristics of the human interacting with the technology. For instance, people with rather social motives in using conversational AI (cf., Choi & Drumwright, 2021) might be generally more likely to perceive the interaction as reciprocal and, at the same time, more willing to disclose personal information when interacting with these systems. Furthermore, less technology-savvy people (e.g., with a low affinity for technology interaction; Franke et al., 2019) might be more likely to ascribe the technology high human-like competencies as a strategy to predict its actions (cf. anthropomorphism as a strategy to predict actions of technology; Epley et al., 2007) as they do not have an idea of the underlying technical process; at the same time these people might be more reluctant to technology use in general and disclosure in specific. Hence, experimental research is needed to validate the causality of the relationships observed in these chapters and rule out that third variables (as described before) drive the observed effects.

Third, there are limitations to the experimental studies conducted within this thesis. It cannot be entirely ruled out that the findings of these studies are limited to the used stimulus materials. This applies in particular to the manipulation of provider trustworthiness in Chapter 4 and the scenario descriptions used in Chapter 5 (both elaborated in the respective chapter discussions). Chapter 2 used more diverse material to manipulate conversational AI competencies, which makes this limitation less likely for the studies reported therein. Furthermore, the vignette designs in Chapters 2 and 5 did not allow for a realistic interaction experience with the

technology within the study, limiting the external validity of the results. Therefore, further studies should ensure that the observed effects can be replicated (a) with other stimuli for manipulation and (b) within more realistic settings.

A Human-Oriented Approach to Investigating Disclosure in Human-AI Interactions

In order to gain a better understanding of disclosure in human-AI interactions, this dissertation took a human-oriented approach building on the well-established ideas of CASA (Nass & Moon, 2000; Reeves & Nass, 1996) and anthropomorphism (Epley et al., 2007). This approach is based on the assumption that humans' interactions with technologies and other humans are – at least to some extent – comparable, as people apply scripts from human-human interactions when interacting with technology (Nass & Moon, 2000; Reeves & Nass, 1996) and ascribe technologies human-like characteristics (Epley et al., 2007). Accordingly, the human-oriented approach focuses on how humans perceive technology rather than on the technical details of implementation – thus making it reasonable to transfer theories, concepts, ideas, and empirical findings from human-human interactions to human-AI interactions.

The overarching framework of this dissertation proposes, in line with findings from human-human interaction, that perceptions of (1) the *interaction partner* (e.g., Boon & Miller, 1999; Collins & Miller, 1994; Qiu et al., 2022) and (2) the *interaction* (e.g., Green et al., 2016; McLaughlin & Cody, 1982; Ramirez et al., 2010; Sunnafrank, 1988; Walsh et al., 2020) impact disclosure in human-AI interactions, which might in turn be related to the acceptance of AI technologies. The results of the present dissertation underline the suitability of this framework. As elaborated in the general introduction of this dissertation, perceptions of an *interaction partner* on a task-related and a socio-emotional dimension are central determinants of behavior in human-human interactions (cf. social evaluation theories; for an integrated overview, see Abele et al., 2021). This seems to apply also to the context of human-AI interactions as the current thesis showed that task-related perceptions of the AI (cf. Chapters 2 and 5) as well as socio-emotional perceptions of the provider (cf. Chapter 4) are crucial for disclosure (and acceptance). Besides, characteristics of the *interaction* seem to be relevant not only in human-human interaction (e.g., Shannon & Weaver, 1964; Sunnafrank, 1986) but also in human-AI interactions: perceptions of higher interactivity (cf. Chapter 3) and higher output quality (cf. Chapter 4) were positively related to disclosure intentions.

While the current dissertation sheds light on the relevance of the aforementioned characteristics for disclosure towards (and acceptance of) AI technologies, many other potentially crucial characteristics could not be addressed within this dissertation. To name only a few, drawing on research from human-human interaction (including computer-mediated

communication), the intimacy (or depth) of information to be disclosed within an *interaction* (e.g., Frye & Dornisch, 2010), or the relationship between the user and other *interaction partners* (e.g., Villalobos Solís et al., 2015), might be crucial for disclosure decisions in human-AI interactions and should be investigated more thoroughly within this framework. Hence, the current dissertation and its overarching framework provide a foundation and a starting point for further research investigating the role of characteristics of the interaction and the interaction partner for disclosure in human-AI interactions following a human-oriented approach.

Technology as a Unique Type of Communicator

However, while humans' interactions with technologies share many similarities with interactions between humans, technologies still constitute a unique type of communicator. Thus, research should not only focus on similarities between humans and technologies but also on the obviously existing differences (cf. Guzman & Lewis, 2020), and, accordingly, must find a balance between human- and technology-oriented approaches.

One research area that considers the specific characteristics of technologies is Human-Computer Interaction (HCI, cf. Excursion 2 in the General Introduction of this dissertation). HCI research is more strongly focused on the physical reality of technologies – including technical characteristics and design features. Accordingly, it can more easily give clear, practical advice on how to design and implement technologies to reach desirable outcomes, such as improved usability and user experience, and, thus, enable optimized human-technology interactions (Guzman, 2018; Vollrath, 2017). As such, it focuses more on technical specificities and less on humans' perception of the interaction partner or the interaction process. The latter perspective is, however, also highly important for understanding human-AI interactions, as demonstrated by the effects observed in the current dissertation. When people's perception of AI differs from the physical reality, their subjective perception rather than the objective technical implementation likely determines how they interact with the system.

Thus, future research should continue to investigate human-AI interactions from different perspectives – including more human-oriented as well as more technology-oriented ones – and integrate these findings in order to get a thorough understanding of how people interact with AI technology. In doing so, the advantages of both approaches can be combined to understand how humans perceive interactions with technology, how they behave in such interactions, and how systems need to be designed to meet the users' needs.

Will Human-AI Interactions Shape Future Human-Human Interactions?

The human-oriented approach in this dissertation is built on the idea that people apply scripts from human-human interaction to human-AI interactions (cf. Nass & Moon, 2000; Reeves & Nass, 1996). However, as people interact with AI more frequently and those scripts might not always be well-applicable, people may also develop own scripts for the interaction with AI (Gambino et al., 2020). In this case, people could apply scripts initially developed for human-AI interaction to human-human interaction. Thus, how we interact with technologies might impact our interpersonal relationships (Gambino et al., 2020). As has been pointed out by several researchers, these processes do not always have to be desirable. For example, people could get used to being less polite in social interactions (Biele et al., 2019) or show stronger tendencies to dehumanize other humans (Waytz et al., 2010). More related to the current dissertation topic, people may also adjust their disclosing behavior toward other humans based on their experiences within the interactions with technologies. For instance, when they perceive that information they disclosed to a technology was used for other than the intended and expected purposes, they might become more reluctant to disclose information in general – also to other humans – which could have severe impacts on personal relationships as disclosure is crucial for building and maintaining relationships between humans (Willems et al., 2020). Hence, while the idea of script transfer from human-AI to human-human interactions goes beyond the scope of the current dissertation, it constitutes a promising avenue for future research, also in the context of disclosure.

Taken together, the effects observed within the current dissertation underline that a human-oriented approach is valuable as it allows a deeper understanding of people's perceptions of and behavior in human-AI interactions. Thus, how humans perceive (AI) technologies should be considered in research as well as in practice. As pointed out above, future studies can further extend the insights of this dissertation by considering (1) additional characteristics of the interaction partner and the interaction from a human-oriented perspective, (2) technology as a unique type of communicator, and (3) potential consequences of human-AI interactions for human-human interactions.

Potential Detrimental Consequences of AI Capabilities

Besides demonstrating the general usefulness of a human-oriented approach to better understand human-AI interactions (elaborated above), the findings of the current dissertation, and Chapters 2 and 5 in particular, highlight one important finding: when perceptions of higher task-related capabilities in an AI surpass a critical threshold, they might no longer evoke positive but instead adverse reactions. This finding resembles the idea of an uncanny valley, which

assumes that more human-likeness (originally in terms of appearance) is – up to a certain point – perceived as beneficial, but if technologies become too human-like, feelings of eeriness can be evoked (Mori et al., 2012; for a review of empirical research, see Kätsyri et al., 2015). Stronger related to the findings of the current thesis, an uncanny valley of mind (Appel et al., 2020; Stein et al., 2020) has recently been proposed: human-like minds in technologies (compared to systems based on basic algorithms) seem to evoke more feelings of eeriness as well (cf. Appel et al., 2020; Stein et al., 2020). Accordingly, not only human-like appearances but also (perceived) task-related capabilities might elicit adverse reactions, such as lower acceptance and less disclosure, if they surpass a critical threshold. A potential explanation for such adverse reactions towards higher AI capabilities is the perceived *agency* of technology (as will be elaborated below). Agency refers to an entity's capacity to exert self-control and, thus, to act and do autonomously (cf. K. Gray & Wegner, 2012). Perceiving agency – and, accordingly, autonomy – in technology can be perceived as a threat to human control (e.g., Stein et al., 2019; Złotowski et al., 2017), which might overshadow potential benefits and result in adverse reactions.

This proposition might explain the adverse reactions to higher AI capabilities observed in the current dissertation. The results of Chapter 2 indicate that perceiving higher competencies on the level of meta-cognitive heuristics reduces disclosure (i.e., heightened privacy concerns; reduced willingness to disclose for non-users) towards conversational AI. Meta-cognitive heuristics refer to the use of abstract and less task-oriented strategies (Zacher, 2017). This implies that the technology 'thinks' (e.g., by transferring 'knowledge' from one domain to another) and 'learns' (e.g., based on previous mistakes and interactions) – suggesting that the technology's actions are no longer under human control but the technology possesses agency itself. In addition, Chapter 5 stresses that managers' AI acceptance (i.e., willingness to invest in and use AI) decreases if the AI system's task-related capabilities (i.e., functionalities) include consequential implementation without a-priori human validation. Such an autonomously implementing technology would obviously also reduce humans' control over the situation and the technology's actions. Thus, the observed adverse reactions towards higher AI capabilities can potentially be explained by the perception of reduced human control caused by technology's agency. Besides, these findings extend the scarce empirical evidence that higher task-related capabilities of technology can evoke adverse reactions (McKee et al., 2022; Złotowski et al., 2017), including heightened privacy concerns (S. X. Liu et al., 2021).

In contrast to the previously described task-related AI capabilities, which were observed to evoke adverse reactions, technology which is capable of dealing with incomplete

information, considering potentially arising problems, or solving unknown tasks (i.e., showing intellectual competencies) is most likely not perceived to possess agency in the narrower sense (i.e., acting autonomously and exerting self-control) and does, therefore, not necessarily limit humans' control. Accordingly, users' positive reactions (i.e., higher willingness to disclose and lower privacy concerns) towards higher task-related capabilities in terms of intellectual competencies (observed in Chapter 2) do not contradict the proposition that perceptions of agency (and the associated risk of losing control) might be the aspect of higher AI capabilities which evokes adverse reactions. However, the finding supports a positive relationship between such task-related technology characteristics and (1) acceptance, as suggested by theoretical models (e.g., TAM; Davis et al., 1989; UTAUT; Venkatesh et al., 2003) and ample empirical evidence (e.g., Dehghani, 2018; Demeure et al., 2011; X. S. Liu et al., 2022; McLean & Osei-Frimpong, 2019; Moriuchi, 2019; Moussawi et al., 2022; Pitardi & Marriott, 2021; Shao & Kwon, 2021), as well as (2) disclosure (e.g., Chellappa & Sin, 2005; Sharma & Crossler, 2014; Townsend et al., 2011; Xu et al., 2013). Thus, the perceived agency of a technology and the associated loss of control for the human might constitute the crucial factor that causes adverse instead of positive reactions (e.g., privacy concerns and lower acceptance) towards highly capable AI. While there is some research demonstrating adverse reactions due to higher autonomy of technologies (e.g., Złotowski et al., 2017), this explanation is, at this point, speculative and needs to be tested systematically in future research.

In addition, future studies should investigate whether similar effects of different capability levels are also observable for other outcomes – such as actual technology usage or performance outcomes in human-AI collaboration. Furthermore, it has to be considered that perceptions of technology and, accordingly, also humans' reactions towards it, could change over time. For instance, it has recently been demonstrated that humans no longer seem to react socially towards desktop computers (Heyselaar, 2023) as assumed in the CASA paradigm (Nass & Moon, 2000; Reeves & Nass, 1996). The authors argue that this change in behavior is based on more interaction experience and familiarization with technologies offering social cues. Similarly, people might also get used to certain task-related capabilities in technology and, as a result, be more willing to accept these capabilities – especially, if experiencing that they still have control while interacting with such technologies. Thus, future research should address whether perceptions of and behavior towards technologies with high task-related capabilities change as soon as people gain interaction experience with these systems and get accustomed to AI capabilities.

Given that higher AI capabilities come with enormous potential – for instance, to enhance comfort and convenience, offer better-personalized services, increase productivity and efficiency, reduce costs, or enable more objective and data-driven decisions (e.g., Brynjolfsson et al., 2018; Hang & Chen, 2022; Sundar, 2020) – it seems imprudent to call for less (or at least not more) capable AI technologies in order to avoid adverse reactions. Instead, we need to gain a better understanding of (1) why high AI capabilities evoke adverse reactions, (2) how these reactions can be avoided (as far as this is desirable), (3) whether reactions towards highly capable AI change over time, and, most importantly, (4) how we can ensure the safe usage of AI systems that harnesses the full potential of AI while minimizing the risks. In this context, besides the previously elaborated perception of agency in technology and the potential loss of control for the human, also other perceptions of threat (e.g., fear of job losses; Broadbent et al., 2012; or human uniqueness concerns; Stein et al., 2019) and a lack of understanding (intransparent) technology (e.g., AI as black box; Castelvecchi, 2016) might constitute crucial factors that should be further investigated.

In addition, the findings regarding AI capabilities have several practical implications. For technology providers, the current findings imply that they should take care of informing (potential) users about what their systems are actually capable of doing – without exaggerating or understating – to reach desirable outcomes such as acceptance and disclosure. This approach should further benefit their trustworthiness (an essential characteristic of the provider as interaction partner, cf. Chapter 4). Moreover, regulators and individuals (as potential users) should be aware that perceptions of AI capabilities impact users' interaction behavior (such as disclosure and usage intentions). Hence, measures need to be taken to ensure that users get a realistic impression of what (AI) technologies can (and will) do (similar to the concept of calibrated trust; cf. J. D. Lee & See, 2004) – for instance, regulating that providers need to be transparent about their systems' capabilities, or educating individuals to have sufficient literacy to make conscious decisions and act safely in human-AI interactions.

Conclusion

Interactions with AI technologies are becoming increasingly common in our everyday lives. Accordingly, almost everyone will interact – and share personal data – with such technologies more or less frequently. To address this issue, this dissertation aimed at a better understanding of disclosure in human-AI interactions. Using a versatile approach, it demonstrates the importance of perceptions of (1) the interaction partner (i.e., the technology and the provider) and (2) the interaction (i.e., interactivity and output quality) in human-AI interactions. As such, it highlights the value of a human-oriented approach that focuses explicitly on how

(potential) users and decision-makers perceive human-AI interactions rather than on the pure technical characteristics of technologies. Furthermore, the empirical results indicate that highly capable AI systems cannot only evoke positive but also adverse reactions (i.e., less disclosure and lower acceptance). Accordingly, future research needs to address how we can ensure safe and beneficial human-AI interactions that harness the full potential of this technology while avoiding adverse reactions and potential risks.

References

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290–314. https://doi.org/10.1037/rev0000262

Alge, B. J., Ballinger, G. A., Tangirala, S., & Oakley, J. L. (2006). Information privacy in organizations: Empowering creative and extrarole performance. *Journal of Applied Psychology*, *91*(1), 221–232. https://doi.org/10.1037/0021-9010.91.1.221

Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior*, *102*, 274–286. https://doi.org/10.1016/j.chb.2019.07.031

Balfe, N., Sharples, S., & Wilson, J. R. (2015). Impact of automation: Measurement of performance, workload and behaviour in a complex control environment. *Applied Ergonomics*, *47*, 52–64. https://doi.org/10.1016/j.apergo.2014.08.002

Balliet, D., & van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112. https://doi.org/10.1037/a0030939

Bandara, R., Fernando, M., & Akter, S. (2020). Privacy concerns in E-commerce: A taxonomy and a future research agenda. *Electronic Markets*, *30*(3), 629–647. https://doi.org/10.1007/s12525-019-00375-6

Bankins, S., & Formosa, P. (2020). When AI meets PC: Exploring the implications of workplace social robots and a human-robot psychological contract. *European Journal of Work and Organizational Psychology*, *29*(2), 215–229. https://doi.org/10.1080/1359432X.2019.1620328

Bansal, G., Zahedi, F. M., & Gefen, D. (2010). The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems*, *49*(2), 138–150. https://doi.org/10.1016/j.dss.2010.01.010

Barth, S., De Jong, M. D. T., Junger, M., Hartel, P. H., & Roppelt, J. C. (2019). Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources. *Telematics and Informatics*, *41*, 55–69. https://doi.org/10.1016/j.tele.2019.03.003

Baruh, L., & Cemalcılar, Z. (2014). It is more than personal: Development and validation of a multidimensional privacy orientation scale. *Personality and Individual Differences*, *70*, 165–170. https://doi.org/10.1016/j.paid.2014.06.042

Baruh, L., Secinti, E., & Cemalcilar, Z. (2017). Online Privacy Concerns and Privacy Management: A Meta-Analytical Review: Privacy Concerns Meta-Analysis. *Journal of Communication*, *67*(1), 26–53. https://doi.org/10.1111/jcom.12276

Beldad, A., & Kusumadewi, M. C. (2015). Here's my location, for your information: The impact of trust, benefits, and social influence on location sharing application use among Indonesian university students. *Computers in Human Behavior*, *49*, 102–110. https://doi.org/10.1016/j.chb.2015.02.047

Beldad, A., Van Der Geest, T., De Jong, M., & Steehouder, M. (2012). Shall I Tell You Where I Live and Who I Am? Factors Influencing the Behavioral Intention to Disclose Personal Data for Online Government Transactions. *International Journal of Human-Computer Interaction*, *28*(3), 163–177. https://doi.org/10.1080/10447318.2011.572331

Beresford, A. R., Kübler, D., & Preibusch, S. (2012). Unwillingness to Pay for Privacy: A Field Experiment. *Economics Letters*, *117*(1), 25–27. https://doi.org/10.1016/j.econlet.2012.04.077

Biele, C., Jaskulska, A., Kopec, W., Kowalski, J., Skorupska, K., & Zdrodowska, A. (2019). How Might Voice Assistants Raise Our Children? In W. Karwowski & T. Ahram (Eds.), *Intelligent Human Systems Integration 2019* (Vol. 903, pp. 162–167). Springer International Publishing. https://doi.org/10.1007/978-3-030-11051-2_25

Bigman, Y., Gray, K., Waytz, A., Arnestad, M., & Wilson, D. (2020). Algorithmic Discrimination Causes Less Moral Outrage than Human Discrimination. *Journal of Experimental Psychology: General*. https://doi.org/10.31234/osf.io/m3nrp

Bleich, H. (2018). Alexa, who has access to my data? Amazon reveals private voice data files. *Heise Online*. https://www.heise.de/downloads/18/2/5/6/5/3/9/6/ct.0119.016-018_engl.pdf

Bol, N., Dienlin, T., Kruikemeier, S., Sax, M., Boerman, S. C., Strycharz, J., Helberger, N., & De Vreese, C. H. (2018). Understanding the Effects of Personalization as a Privacy Calculus: Analyzing Self-Disclosure Across Health, News, and Commerce Contexts. *Journal of Computer-Mediated Communication*, *23*(6), 370–388. https://doi.org/10.1093/jcmc/zmy020

Boon, S. D., & Miller, R. J. (1999). Exploring the Links Between Interpersonal Trust and the Reasons Underlying Gay and Bisexual Males' Disclosure of Their Sexual Orientation to Their Mothers. *Journal of Homosexuality*, *37*(3), 45–68. https://doi.org/10.1300/J082v37n03_04

Boyles, J. L., Smith, A., & Madden, M. (2012). *Privacy and Data Management on Mobile Devices* (4; Pew Internet & American Life Project, pp. 1–19).

Broadbent, E., Tamagawa, R., Patience, A., Knock, B., Kerse, N., Day, K., & MacDonald, B. A. (2012). Attitudes towards health-care robots in a retirement village. *Australasian Journal on Ageing*, *31*(2), 115–120. https://doi.org/10.1111/j.1741-6612.2011.00551.x

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530–1534. https://doi.org/10.1126/science.aap8062

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What Can Machines Learn and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, *108*, 43–47. https://doi.org/10.1257/pandp.20181019

Burgoon, J. K. (1982). Privacy and communication. In *Communication Yearbooks* (1st ed., Vol. 6, pp. 206–249). Routledge.

Campbell, D. (2019). A Relational Build-up Model of Consumer Intention to Self-disclose Personal Information in E-commerce B2C Relationships. *AIS Transactions on Human-Computer Interaction*, *11*(1), 33–53. https://doi.org/10.17705/1thci.00112

Carrascal, J. P., Riederer, C., Erramilli, V., Cherubini, M., & De Oliveira, R. (2013). Your browsing behavior for a big mac: Economics of personal information online. *Proceedings of the 22nd International Conference on World Wide Web*, 189–200. https://doi.org/10.1145/2488388.2488406

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, *538*(7623), 20–23. https://doi.org/10.1038/538020a

Chellappa, R. K., & Sin, R. G. (2005). Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. *Information Technology and Management*, *6*, 181–202. https://doi.org/10.1007/s10799-005-5879-y

Chen, L., Zarifis, A., & Kroenung, J. (2017). The role of trust in personal information disclosure on health-related websites. *Proceedings of the European Conference on Information Systems (ECIS)*, *1*, 771–786. http://aisel.aisnet.org/ecis2017_rp/50

Cheng, Y.-M. (2014). Roles of interactivity and usage experience in e-learning acceptance: A longitudinal study. *International Journal of Web Information Systems*, *10*(1), 2–23. https://doi.org/10.1108/IJWIS-05-2013-0015

Choi, T. R., & Drumwright, M. E. (2021). “OK, Google, why do I use you?” Motivations, post-consumption evaluations, and perceptions of voice AI assistants. *Telematics and Informatics*, *62*, 101628. https://doi.org/10.1016/j.tele.2021.101628

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, *116*(3), 457–475. https://doi.org/10.1037/0033-2909.116.3.457

Confessore, N. (2018, April 4). *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*. The New York Times. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Copeland, B. J. (2023, August 9). *Artificial intelligence (AI)*. Encyclopaedia Britannica. https://www.britannica.com/technology/artificial-intelligence

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*(4), 631–648. https://doi.org/10.1037/0022-3514.92.4.631

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. In *Advances in Experimental Social Psychology* (Vol. 40, pp. 61–149). Elsevier. https://doi.org/10.1016/S0065-2601(07)00002-0

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

de Graaf, M. M. A., & Ben Allouch, S. (2013). Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, *61*(12), 1476–1486. https://doi.org/10.1016/j.robot.2013.07.007

Dehghani, M. (2018). Exploring the motivational factors on continuous usage intention of smartwatches among actual users. *Behaviour & Information Technology*, *37*(2), 145–158. https://doi.org/10.1080/0144929X.2018.1424246

Demeure, V., Niewiadomski, R., & Pelachaud, C. (2011). How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment? *Presence: Teleoperators and Virtual Environments*, *20*(5), 431–448. https://doi.org/10.1162/PRES_a_00065

Dienlin, T., & Metzger, M. J. (2016). An Extended Privacy Calculus Model for SNSs: Analyzing Self-Disclosure and Self-Withdrawal in a Representative U.S. Sample. *Journal of Computer-Mediated Communication*, *21*(5), 368–383. https://doi.org/10.1111/jcc4.12163

Dineen, B. R., Noe, R. A., & Wang, C. (2004). Perceived fairness of web-based applicant screening procedures: Weighing the rules of justice and the role of individual differences. *Human Resource Management*, *43*(2–3), 127–145. https://doi.org/10.1002/hrm.20011

Dinev, T., & Hart, P. (2006). An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research*, *17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

D'Souza, G., & Phelps, J. E. (2009). The Privacy Paradox: The Case of Secondary Disclosure. *Review of Marketing Science*, *7*, 4. https://doi.org/10.2202/1546-5616.1072

Duarte, F. (2023, July 13). *Number of ChatGPT Users in 2023*. Exploding Topics. https://explodingtopics.com/blog/chatgpt-users

Dubiel, M., Halvey, M., & Azzopardi, L. (2018). *A Survey Investigating Usage of Virtual Personal Assistants* (arXiv:1807.04606). arXiv. http://arxiv.org/abs/1807.04606

Easwara Moorthy, A., & Vu, K.-P. L. (2015). Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human-Computer Interaction*, *31*(4), 307–335. https://doi.org/10.1080/10447318.2014.986642

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

European Union. (2018, May 15). *General Data Protection Regulation (GDPR)*. https://gdpr.eu/

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating Affordances: A Conceptual Framework for Understanding Affordances in Communication Research. *Journal of Computer-Mediated Communication*, *22*(1), 35–52. https://doi.org/10.1111/jcc4.12180

Fan, L., Liu, X., Wang, B., & Wang, L. (2017). Interactivity, engagement, and technology dependence: Understanding users' technology utilisation behaviour. *Behaviour & Information Technology*, *36*(2), 113–124. https://doi.org/10.1080/0144929X.2016.1199051

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, *25*(1), 153–160. https://doi.org/10.1016/j.chb.2008.08.006

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

Frese, M., & Zapf, D. (1994). Action as the core of work psychology A German approach. In H. C. Trianidis, M. D. Dunette, & L. M. Hough (Eds.), *Handbook of industrial and orgnizational psychology* (Vol. 4, pp. 271–340).

Frye, N. E., & Dornisch, M. M. (2010). When is trust not enough? The role of perceived privacy of communication tools in comfort with self-disclosure. *Computers in Human Behavior*, *26*(5), 1120–1127. https://doi.org/10.1016/j.chb.2010.03.016

Galière, S. (2020). When food-delivery platform workers consent to algorithmic management: A Foucauldian perspective. *New Technology, Work and Employment*, *35*(3), 357–370. https://doi.org/10.1111/ntwe.12177

Gambino, A., Fox, J., & Ratan, R. (2020). Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. *Human-Machine Communication*, *1*, 71–86. https://doi.org/10.30658/hmc.1.5

Gerlach, J., Widjaja, T., & Buxmann, P. (2015). Handle with care: How online social network providers' privacy policies impact users' information sharing behavior. *The Journal of Strategic Information Systems*, *24*(1), 33–43. https://doi.org/10.1016/j.jsis.2014.09.001

Gieselmann, M., & Sassenberg, K. (2022). Dataset for: The More Competent, the Better? The Effects of Perceived Competencies on Disclosure Towards Conversational Artificial Intelligence. *PsychArchives*. https://doi.org/10.23668/PSYCHARCHIVES.12175

Gieselmann, M., & Sassenberg, K. (2023a). Code for: The Relevance of Perceived Interactivity for Disclosure Towards Conversational Artificial Intelligence. *PsychArchives*. https://doi.org/10.23668/PSYCHARCHIVES.12511

Gieselmann, M., & Sassenberg, K. (2023b). The More Competent, the Better? The Effects of Perceived Competencies on Disclosure Towards Conversational Artificial Intelligence. *Social Science Computer Review*, *41*(6), 2342–2363. https://doi.org/10.1177/08944393221142787

Gikopoulos, J. (2019). Alongside, not against: Balancing man with machine in the HR function. *Strategic HR Review*, *18*(2), 56–61. https://doi.org/10.1108/SHR-12-2018-0103

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of Mind Perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/science.1134475

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Green, T., Wilhelmsen, T., Wilmots, E., Dodd, B., & Quinn, S. (2016). Social anxiety, attributes of online communication and self-disclosure across private and public Facebook communication. *Computers in Human Behavior*, *58*, 206–213. https://doi.org/10.1016/j.chb.2015.12.066

Greussing, E., Gaiser, F., Klein, S. H., Straßmann, C., Ischen, C., Eimler, S., Frehmann, K., Gieselmann, M., Knorr, C., Lermann Henestrosa, A., Räder, A., & Utz, S. (2022). Researching interactions between humans and machines: Methodological challenges. *Publizistik*, *67*(4), 531–554. https://doi.org/10.1007/s11616-022-00759-3

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Guzman, A. L. (2018). What is human-machine communication, anyway? In *Human-machine communication: Rethinking communication, technology and ourselves* (pp. 1–28). Peter Lang.

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

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society*, *22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Ha, Q.-A., Chen, J. V., Uy, H. U., & Capistrano, E. P. (2021). Exploring the Privacy Concerns in Using Intelligent Virtual Assistants under Perspectives of Information Sensitivity and Anthropomorphism. *International Journal of Human–Computer Interaction*, *37*(6), 512–527. https://doi.org/10.1080/10447318.2020.1834728

Hacker, W. (1971). *Allgemeine Arbeits- und Ingenieurpsychologie* [General work and engineering psychology]. Deutscher Verlag der Wissenschaften.

Hacker, W. (1998). *Allgemeine Arbeitspsychologie: Psychische Regulation von Arbeitstätigkeiten* [General work psychology: Mental regulation of work tasks]. Huber.

Hacker, W., & Sachse, P. (2014). *Allgemeine Arbeitspsychologie* (3rd ed.) [General work psychology]. Hogrefe.

Hang, H., & Chen, Z. (2022). How to realize the full potentials of artificial intelligence (AI) in digital economy? A literature review. *Journal of Digital Economy*, *1*(3), 180–191. https://doi.org/10.1016/j.jdec.2022.11.003

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future? *AI*, *1*(2), 143–155. https://doi.org/10.3390/ai1020008

Heeter, C. (2000). Interactivity in the Context of Designed Experiences. *Journal of Interactive Advertising*, *1*(1), 3–14. https://doi.org/10.1080/15252019.2000.10722040

Hepp, A., Loosen, W., Dreyer, S., Jarke, J., Kannengießer, S., Katzenbach, C., Malaka, R., Pfadenhauer, M., Puschmann, C., & Schulz, W. (2022). Von der Mensch-Maschine-Interaktion zur kommunikativen KI: Automatisierung von Kommunikation als Gegenstand der Kommunikations- und Medienforschung [From Human-Machine Interaction to Communicative AI: Automation of Communication as a Subject of Communication and Media Research]. *Publizistik*, *67*(4), 449–474. https://doi.org/10.1007/s11616-022-00758-4

Heyselaar, E. (2023). The CASA theory no longer applies to desktop computers. *Scientific Reports*, *13*(1), 19693. https://doi.org/10.1038/s41598-023-46527-9

Höddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, *116*, 106635. https://doi.org/10.1016/j.chb.2020.106635

Hong, J.-W., Choi, S., & Williams, D. (2020). Sexist AI: An Experiment Integrating CASA and ELM. *International Journal of Human–Computer Interaction*, *36*(20), 1928–1941. https://doi.org/10.1080/10447318.2020.1801226

Horstmann, A., Szczuka, J., Mavrina, L., Artelt, A., Strathmann, C., Szymczyk, N., Bohnenkamp, L. M., & Krämer, N. (2023, May 25). *Enhancing the Understanding of Algorithms With Contrastive Explanations: An Experimental Study on the Effects of Explanations and Person-Likeness on Trust in and Understanding of Algorithms*. 73rd Annual International Communication Association Conference, Toronto, Canada.

Johnson, T. J., & Kaye, B. K. (2016). Some like it lots: The influence of interactivity and reliance on credibility. *Computers in Human Behavior*, *61*, 136–145. https://doi.org/10.1016/j.chb.2016.03.012

Joinson, A., Reips, U.-D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, Trust, and Self-Disclosure Online. *Human-Computer Interaction*, *25*(1), 1–24. https://doi.org/10.1080/07370020903586662

Jutzi, T. B., Krieghoff-Henning, E. I., Holland-Letz, T., Utikal, J. S., Hauschild, A., Schadendorf, D., Sondermann, W., Fröhling, S., Hekler, A., Schmitt, M., Maron, R. C., & Brinker, T. J. (2020). Artificial Intelligence in Skin Cancer Diagnostics: The Patients' Perspective. *Frontiers in Medicine*, *7*, 233. https://doi.org/10.3389/fmed.2020.00233

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, *5*(2), 113–153. https://doi.org/10.1080/1463922021000054335

Kalman, Y. M., Scissors, L. E., & Gergle, D. (2010). Chronemic aspects of chat, and their relationship to trust in a virtual team. *MCIS 2010 Proceedings*, *46*. https://aisel.aisnet.org/mcis2010/46

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00390

Keding, C., & Meissner, P. (2021). Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions. *Technological Forecasting and Social Change*, *171*, 120970. https://doi.org/10.1016/j.techfore.2021.120970

Kezer, M., Dienlin, T., & Baruh, L. (2022). Getting the privacy calculus right: Analyzing the relations between privacy concerns, expected benefits, and self-disclosure using response surface analysis. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *16*(4). https://doi.org/10.5817/CP2022-4-1

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, *28*(1), 241–250. https://doi.org/10.1016/j.chb.2011.09.006

Kim-Schmid, J., & Raveendhran, R. (2022, October 13). *Where AI Can—And Can't—Help Talent Management*. https://hbr.org/2022/10/where-ai-can-and-cant-help-talent-management

Kinsella, B. (2021, April 14). *U.S. Smart Speaker Growth Flat Lined in 2020*. Voicebot.Ai. https://voicebot.ai/2021/04/14/u-s-smart-speaker-growth-flat-lined-in-2020/

Kinsella, B. (2023, June 17). *Germany Smart Speaker Adoption Closely Mirrors U.S. Pattern—New Report with 30+ Charts*. Voicebot.Ai. https://voicebot.ai/2021/06/17/germany-smart-speaker-adoption-closely-mirrors-u-s-pattern-new-report-with-30-charts/

Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*, 106553. https://doi.org/10.1016/j.chb.2020.106553

Kok, J. N., Boers, E. J., Kosters, W A, Van der Puten, P., & Poel, M. (2009). Artificial Intelligence: Definition, Trends, Techniques and Cases. *Artificial Intelligence*, *1*, 270–299.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Krapinger, G. (Ed.). (2018). *Aristoteles Rhetorik*. Reclam.

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online Social Networks: Why We Disclose. *Journal of Information Technology*, *25*(2), 109–125. https://doi.org/10.1057/jit.2010.6

Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, *27*(3), 217–234. https://doi.org/10.1111/ijsa.12246

Langer, M., König, C. J., Sanchez, D. R.-P., & Samadi, S. (2019). Highly automated interviews: Applicant reactions and the organizational context. *Journal of Managerial Psychology*, *35*(4), 301–314. https://doi.org/10.1108/JMP-09-2018-0402

Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, *123*, 106878. https://doi.org/10.1016/j.chb.2021.106878

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What Do We Want From Explainable Artificial Intelligence (XAI)? A Stakeholder Perspective on XAI and a Conceptual Model Guiding Interdisciplinary XAI

Research. *Artificial Intelligence*, *296*, 103473. https://doi.org/10.1016/j.artint.2021.103473

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*.

Lee, J.-H., & Song, C.-H. (2013). Effects of trust and perceived risk on user acceptance of a new technology service. *Social Behavior and Personality: An International Journal*, *41*(4), 587–597. https://doi.org/10.2224/sbp.2013.41.4.587

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718756684

Lew, Z., & Stohl, C. (2022). What makes people willing to comment on social media posts? The roles of interactivity and perceived contingency in online corporate social responsibility communication. *Communication Monographs*, *90*(1), 1–24. https://doi.org/10.1080/03637751.2022.2032230

Lew, Z., Walther, J. B., Pang, A., & Shin, W. (2018). Interactivity in Online Chat: Conversational Contingency and Response Latency in Computer-mediated Communication. *Journal of Computer-Mediated Communication*, *23*(4), 201–221. https://doi.org/10.1093/jcmc/zmy009

Liao, Y., Vitak, J., Kumar, P., Zimmer, M., & Kritikos, K. (2019). Understanding the Role of Privacy and Trust in Intelligent Personal Assistant Adoption. In N. G. Taylor, C. Christian-Lamb, M. H. Martin, & B. Nardi (Eds.), *LNCS: Information in Contemporary Society* (Vol. 11420, pp. 102–113). Springer International Publishing. https://doi.org/10.1007/978-3-030-15742-5_9

Liu, S. X., Shen, Q., & Hancock, J. (2021). Can a social robot be too warm or too competent? Older Chinese adults' perceptions of social robots and vulnerabilities. *Computers in Human Behavior*, *125*, 106942. https://doi.org/10.1016/j.chb.2021.106942

Liu, X., Min, Q., & Han, S. (2020). Understanding users' continuous content contribution behaviours on microblogs: An integrated perspective of uses and gratification theory and social influence theory. *Behaviour & Information Technology*, *39*(5), 28. https://doi.org/10.1080/0144929X.2019.1603326

Liu, X. S., Yi, X. S., & Wan, L. C. (2022). Friendly or competent? The effects of perception of robot appearance and service context on usage intention. *Annals of Tourism Research*, *92*, 103324. https://doi.org/10.1016/j.annals.2021.103324

Liu, Y. (2003). Developing a scale to measure the interactivity of websites. *Journal of Advertising Research*, *43*(2), 207–216. https://doi.org/10.2501/JAR-43-2-207-216

Lombard, M., & Xu, K. (2021). Social Responses to Media Technologies in the 21st Century: The Media are Social Actors Paradigm. *Human-Machine Communication*, *2*, 29–55. https://doi.org/10.30658/hmc.2.2

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Luger, G. F. (2009). *Artificial intelligence: Structures and strategies for complex problem solving* (6th ed). Pearson Addison-Wesley.

Lynskey, D. (2019, October 9). "Alexa, are you invading my privacy?" – the dark side of our voice assistants. *The Guardian*. https://www.theguardian.com/technology/2019/oct/09/alexa-are-you-invading-my-privacy-the-dark-side-of-our-voice-assistants

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, *15*(4), 336–355. https://doi.org/10.1287/isre.1040.0032

Marek, C. I., Wanzer, M. B., & Knapp, J. L. (2004). An exploratory investigation of the relationship between roommates' first impressions and subsequent communication patterns. *Communication Research Reports*, *21*(2), 210–220. https://doi.org/10.1080/08824090409359982

Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016). Tell me more designing HRI to encourage more trust, disclosure, and companionship. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 181–188. https://doi.org/10.1109/HRI.2016.7451750

Mayer, R. C., Davis, J. H., & Schoorman, D. F. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McKee, K. R., Bai, X., & Fiske, S. T. (2022). *Warmth and competence in human-agent cooperation* (arXiv:2201.13448). arXiv. http://arxiv.org/abs/2201.13448

McKnight, D. H., & Chervany, N. L. (2001). What Trust Means in E-Commerce Customer Relationships: An Interdisciplinary Conceptual Typology. *International Journal of Electronic Commerce*, *6*(2), 35–59. https://doi.org/10.1080/10864415.2001.11044235

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*, *13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

McLaughlin, M. L., & Cody, M. J. (1982). Awkward silences: Behavioral Antecedents and consequences of the conversational lapse. *Human Communication Research*, *8*(4), 299–316. https://doi.org/10.1111/j.1468-2958.1982.tb00669.x

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28–37. https://doi.org/10.1016/j.chb.2019.05.009

Mediavision Interactive. (2021, July 30). *The value of your personal data*. MiQ. https://www.wearemiq.com/blog/value-of-personal-data/

Mesch, G. S. (2012). Is online trust and trust in social institutions associated with online disclosure of identifiable information online? *Computers in Human Behavior*, *28*(4), 1471–1477. https://doi.org/10.1016/j.chb.2012.03.010

Metzger, M. J. (2006). Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce. *Journal of Computer-Mediated Communication*, *9*(4), Article JMCM942. https://doi.org/10.1111/j.1083-6101.2004.tb00292.x

Milmo, D. (2023, February 2). *ChatGPT reaches 100 million users two months after launch*. The Guardian. https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app

Milne, G. R., Rohm, A. J., & Bahl, S. (2004). Consumers' Protection of Online Privacy and Identity. *Journal of Consumer Affairs*, *38*(2), 217–232. https://doi.org/10.1111/j.1745-6606.2004.tb00865.x

Miltgen, C. L., & Peyrat-Guillard, D. (2014). Cultural and generational influences on privacy concerns: A qualitative study in seven European countries. *European Journal of Information Systems*, *23*(2), 103–125. https://doi.org/10.1057/ejis.2013.17

Mohamed, N., & Ahmad, I. H. (2012). Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior*, *28*(6), 2366–2375. https://doi.org/10.1016/j.chb.2012.07.008

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. https://doi.org/10.25300/MISQ/2021/15333

Montgomery, A. L., & Smith, M. D. (2009). Prospects for Personalization on the Internet. *Journal of Interactive Marketing*, *23*(2), 130–137. https://doi.org/10.1016/j.intmar.2009.02.001

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Moriuchi, E. (2019). Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. *Psychology & Marketing*, *36*(5), 489–501. https://doi.org/10.1002/mar.21192

Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2021). How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets*, *31*(2), 343–364. https://doi.org/10.1007/s12525-020-00411-w

Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2022). The role of user perceptions of intelligence, anthropomorphism, and self-extension on continuance of use of personal intelligent agents. *European Journal of Information Systems*, *32*(3), 601–622. https://doi.org/10.1080/0960085X.2021.2018365

Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly*, *38*(1), 101536. https://doi.org/10.1016/j.giq.2020.101536

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

Nayyar, S. (2023). Is ChatGPT Really One Of The Most Important Milestones In Digital Technology? *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2023/04/03/is-chatgpt-really-one-of-the-most-important-milestones-in-digital-technology/

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, *160*, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

Nienaber, A.-M., & Schewe, G. (2014). Enhancing trust or reducing perceived risk, what matters more when launching a new product? *International Journal of Innovation Management*, *18*(1), 1450005. https://doi.org/10.1142/S1363919614500054

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs*, *41*(1), 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

Open AI. (2023, August 3). *ChatGPT*. https://chat.openai.com/auth/login

Pal, D., Arpnikanondt, C., & Razzaque, M. A. (2020). Personal Information Disclosure via Voice Assistants: The Personalization–Privacy Paradox. *SN Computer Science*, *1*(5), 280. https://doi.org/10.1007/s42979-020-00287-9

Palmisciano, P., Jamjoom, A. A. B., Taylor, D., Stoyanov, D., & Marcus, H. J. (2020). Attitudes of Patients and Their Relatives Toward Artificial Intelligence in Neurosurgery. *World Neurosurgery*, *138*, e627–e633. https://doi.org/10.1016/j.wneu.2020.03.029

Park, E. K., & Sundar, S. S. (2015). Can synchronicity and visual modality enhance social presence in mobile messaging? *Computers in Human Behavior*, *45*, 121–128. https://doi.org/10.1016/j.chb.2014.12.001

Park, Y. W., & Lee, A. R. (2019). The moderating role of communication contexts: How do media synchronicity and behavioral characteristics of mobile messenger applications affect social intimacy and fatigue? *Computers in Human Behavior*, *97*, 179–192. https://doi.org/10.1016/j.chb.2019.03.020

Peng, Z., Kwon, Y., Lu, J., Wu, Z., & Ma, X. (2019). Design and Evaluation of Service Robot's Proactivity in Decision-Making Support Process. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300328

Pitardi, V., & Marriott, H. R. (2021). Alexa, she's not human but… Unveiling the drivers of consumers' trust in voice-based artificial intelligence. *Psychology & Marketing*, *38*(4), 626–642. https://doi.org/10.1002/mar.21457

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*, 2853–2859. https://doi.org/10.1145/3027063.3053246

Qiu, J., Kesebir, S., Günaydin, G., Selçuk, E., & Wasti, S. A. (2022). Gender differences in interpersonal trust: Disclosure behavior, benevolence sensitivity and workplace implications. *Organizational Behavior and Human Decision Processes*, *169*, 104119. https://doi.org/10.1016/j.obhdp.2022.104119

Ramirez, A., Sunnafrank, M., & Goei, R. (2010). Predicted Outcome Value Theory in Ongoing Relationships. *Communication Monographs*, *77*(1), 27–50. https://doi.org/10.1080/03637750903514276

Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press.

Reynolds, B., Venkatanathan, J., Gonçalves, J., & Kostakos, V. (2011). Sharing Ephemeral Information in Online Social Networks: Privacy Perceptions and Behaviours. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2011* (Vol. 6948, pp. 204–215). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23765-2_14

Rimol, M. (2022, August 22). *Gartner Survey Reveals 80% of Executives Think Automation Can Be Applied to Any Business Decision*. https://www.gartner.com/en/news-room/press-releases/2022-08-22-gartner-survey-reveals-80-percent-of-executives-think-automation-can-be-applied-to-any-business-decision

Rosenthal, S., Wasenden, O.-C., Gronnevet, G.-A., & Ling, R. (2020). A tripartite model of trust in Facebook: Acceptance of information personalization, privacy concern, and privacy literacy. *Media Psychology*, *23*(6), 840–864. https://doi.org/10.1080/15213269.2019.1648218

Rosseel, Y., Jorgensen, T. D., & Rockwood, N. (2021). *Package "lavaan" [R package]* [Computer software]. https://cran.r-project.org/web/packages/lavaan/lavaan.pdf

Schank, R. C., & Abelson, R. P. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures*. Taylor and Francis.

Schoenbachler, D. D., & Gordon, G. L. (2002). Trust and customer willingness to provide information in database-driven relationship marketing. *Journal of Interactive Marketing*, *16*(3), 2–16. https://doi.org/10.1002/dir.10033

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schouten, A. P., Valkenburg, P. M., & Peter, J. (2007). Precursors and Underlying Processes of Adolescents' Online Self-Disclosure: Developing and Testing an "Internet-Attribute-Perception" Model. *Media Psychology*, *10*(2), 292–315. https://doi.org/10.1080/15213260701375686

Semmer, N., & Frese, M. (1985). Action theory in clinical psychology. In M. Frese & J. Sabini (Eds.), *Goal-Directed Behavior The Concept of Action in Psychology* (pp. 296–310). Lawrence Hillbaum.

Shank, D. B., Graves, C., Gott, A., Gamez, P., & Rodriguez, S. (2019). Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence. *Computers in Human Behavior*, *98*, 256–266. https://doi.org/10.1016/j.chb.2019.04.001

Shannon, C., & Weaver, W. (1964). *The Mathematical Theory of Communication* (10th ed.). The University of Illinois Press.

Shao, C., & Kwon, K. H. (2021). Hello Alexa! Exploring effects of motivational factors and social presence on satisfaction with artificial intelligence-enabled gadgets. *Human Behavior and Emerging Technologies*, *3*(5), 978–988. https://doi.org/10.1002/hbe2.293

Sharma, S., & Crossler, R. E. (2014). Disclosing too much? Situational factors affecting information disclosure in social commerce environment. *Electronic Commerce Research and Applications*, *13*(5), 305–319. https://doi.org/10.1016/j.elerap.2014.06.007

Shin, D.-H., Hwang, Y., & Choo, H. (2013). Smart TV: Are they really smart in interacting with people? Understanding the interactivity of Korean Smart TV. *Behaviour & Information Technology*, *32*(2), 156–172. https://doi.org/10.1080/0144929X.2011.603360

Shu, W. (2014). Continual use of microblogs. *Behaviour & Information Technology*, *33*(7), 666–677. https://doi.org/10.1080/0144929X.2013.816774

Spiekermann, S., Großklags, J., & Berendt, B. (2001). Stated Privacy Preferences versus Actual Behaviour in EC Environments: A Reality Check. In H. U. Buhl, N. Kreyer, & W. Steck (Eds.), *E-Finance* (pp. 129–147). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-59504-2_8

Statista. (2022, March 14). *Number of voice assistants in use worldwide 2019-2024*. https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/

Stein, J.-P., Appel, M., Jost, A., & Ohler, P. (2020). Matter over mind? How the acceptance of digital entities depends on their appearance, mental prowess, and the interaction between both. *International Journal of Human-Computer Studies*, *142*, 102463. https://doi.org/10.1016/j.ijhcs.2020.102463

Stein, J.-P., Liebold, B., & Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Computers in Human Behavior*, *95*, 73–82. https://doi.org/10.1016/j.chb.2019.01.021

Suler, J. (2012). The Online Disinhibition Effect. *Cyberpsychology, Behavior and Social Networking*, *15*(2), 103–111. https://doi.org/10.1089/cyber.2011.0277

Sundar, S. S. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital Media, Youth, and Credibility* (pp. 73–100). The MIT Press.

Sundar, S. S. (2020). Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Sunnafrank, M. (1986). Predicted Outcome Value During Initial Interactions A Reformulation of Uncertainty Reduction Theory. *Human Communication Research*, *13*(1), 3–33. https://doi.org/10.1111/j.1468-2958.1986.tb00092.x

Sunnafrank, M. (1988). Predicted outcome value in initial conversations. *Communication Research Reports*, *5*(2), 169–172. https://doi.org/10.1080/08824098809359819

Sunnafrank, M., & Ramirez, A. (2004). At First Sight: Persistent Relational Effects of Get-Acquainted Conversations. *Journal of Social and Personal Relationships*, *21*(3), 361–379. https://doi.org/10.1177/0265407504042837

Swan, J. E., Bowers, M. R., & Richardson, L. D. (1999). Customer Trust in the Salesperson: An Integrative Review and Meta-Analysis of the Empirical Literature. *Journal of Business Research*, *44*(2), 93–107.

Taddei, S., & Contena, B. (2013). Privacy, trust and control: Which relationships with online self-disclosure? *Computers in Human Behavior*, *29*(3), 821–826. https://doi.org/10.1016/j.chb.2012.11.022

Taddicken, M. (2014). The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure. *Journal of Computer-Mediated Communication*, *19*(2), 248–273. https://doi.org/10.1111/jcc4.12052

Tavani, H. T. (2008). Informational Privacy: Concepts, Theories, and Controversies. In K. E. Himma & H. T. Tavani (Eds.), *Handbook of Information and Computer Ethics* (pp. 131–164). Wiley.

Townsend, D., Knoefel, F., & Goubran, R. (2011). Privacy versus autonomy: A tradeoff model for smart home monitoring technologies. *Proceedings of the 2011 Annu Int Conf IEEE Eng Med Biol Soc*, 4749–4752. https://doi.org/10.1109/IEMBS.2011.6091176

Trepte, S., Scharkow, M., & Dienlin, T. (2020). The privacy calculus contextualized: The influence of affordances. *Computers in Human Behavior*, *104*, 106115. https://doi.org/10.1016/j.chb.2019.08.022

Tschopp, M., Scharowski, N., & Wintersberger, P. (2021, June 1). *Do Humans Trust AI or Its Developers? Exploring Benefits of Differentiating Trustees Within Trust in AI Frameworks*. Conference: Workshop: The Culture of trustworthy AI. https://repositum.tuwien.at/handle/20.500.12708/87271?mode=full

Tufekci, Z. (2008). Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites. *Bulletin of Science, Technology & Society*, *28*(1), 20–36. https://doi.org/10.1177/0270467607311484

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, *39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Villalobos Solís, M., Smetana, J. G., & Comer, J. (2015). Associations among solicitation, relationship quality, and adolescents' disclosure and secrecy with mothers and best friends. *Journal of Adolescence*, *43*(1), 193–205. https://doi.org/10.1016/j.adolescence.2015.05.016

Vollrath, M. (2017). *Ingenieurpsychologie: Psychologische Grundlagen und Anwendungsgebiete* [Engineering psychology: Psychological principles and areas of application]. Kohlhammer.

Wakefield, R. (2013). The influence of user affect in online information disclosure. *The Journal of Strategic Information Systems*, *22*(2), 157–174. https://doi.org/10.1016/j.jsis.2013.01.003

Walrave, M., Vanwesenbeeck, I., & Heirman, W. (2012). Connecting and protecting? Comparing predictors of self-disclosure and privacy settings use between adolescents and adults. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *6*(1), Article 3. https://doi.org/10.5817/CP2012-1-3

Walsh, R. M., Forest, A. L., & Orehek, E. (2020). Self-disclosure on social media: The role of perceived network responsiveness. *Computers in Human Behavior*, *104*, 106162. https://doi.org/10.1016/j.chb.2019.106162

Warwick, K., & Shah, H. (2016). Can machines think? A report on Turing test experiments at the Royal Society. *Journal of Experimental & Theoretical Artificial Intelligence*, *28*(6), 989–1007. https://doi.org/10.1080/0952813X.2015.1055826

Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social Cognition Unbound: Insights Into Anthropomorphism and Dehumanization. *Current Directions in Psychological Science*, *19*(1), 58–62. https://doi.org/10.1177/0963721409359302

Westin, A. F. (2003). Social and Political Dimensions of Privacy. *Journal of Social Issues*, *59*(2), 431–453. https://doi.org/10.1111/1540-4560.00072

Wheeless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research*, *3*(3), 250–257. https://doi.org/10.1111/j.1468-2958.1977.tb00523.x

Willems, Y. E., Finkenauer, C., & Kerkhof, P. (2020). The role of disclosure in relationships. *Current Opinion in Psychology*, *31*, 33–37. https://doi.org/10.1016/j.copsyc.2019.07.032

Wottrich, V. M., Van Reijmersdal, E. A., & Smit, E. G. (2018). The privacy trade-off for mobile app downloads: The roles of app value, intrusiveness, and privacy concerns. *Decision Support Systems*, *106*, 44–52. https://doi.org/10.1016/j.dss.2017.12.003

Wottrich, V. M., Verlegh, P. W. J., & Smit, E. G. (2017). The role of customization, brand trust, and privacy concerns in advergaming. *International Journal of Advertising*, *36*(1), 60–81. https://doi.org/10.1080/02650487.2016.1186951

Xu, F., Michael, K., & Chen, X. (2013). Factors affecting privacy disclosure on social network sites: An integrated model. *Electronic Commerce Research*, *13*(2), 151–168. https://doi.org/10.1007/s10660-013-9111-6

Yan, C., Ji, X., Wang, K., Jiang, Q., Jin, Z., & Xu, W. (2022). A Survey on Voice Assistant Security: Attacks and Countermeasures. *ACM Computing Surveys*, *55*(4), 1–36. https://doi.org/10.1145/3527153

Yang, S., & Wang, K. (2009). The influence of information sensitivity compensation on privacy concern and behavioral intention. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *40*(1), 38–51. https://doi.org/10.1145/1496930.1496937

Zacher, H. (2017). Action Regulation Theory. *Oxford Research Encyclopedia of Psychology*. https://doi.org/10.1093/acrefore/9780190236557.013.1

Zimmer, J. C., Arsal, R. E., Al-Marzouq, M., & Grover, V. (2010). Investigating online information disclosure: Effects of information relevance, trust and risk. *Information & Management*, *47*(2), 115–123. https://doi.org/10.1016/j.im.2009.12.003

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, *100*, 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008

Appendix

Appendix A: Additional Information for Chapter 2[4]

Data sets (http://dx.doi.org/10.23668/psycharchives.12175) and analyses syntaxes (http://dx.doi.org/10.23668/psycharchives.12176) for all studies reported in Chapter 2 are available for scientific use via PsychArchives.

Additional Information for Study 1.1

Exclusion of Participants

In total, the survey was started by 417 participants and completed with the consent to use data by 399 participants. One participant was excluded form analysis as they participated twice in the survey. Further, as pre-registered, we excluded a total of 40 participants who (1) did not complete the survey using a computer or tablet, (2) failed two of two attention checks and/or (3) did not consider their own data usable for scientific research. Thus, the eligible sample consisted of $N = 358$ participants (as reported in Chapter 2).

Pre-Registration

Created:	03/15/2021 06:18 AM (PT)
Made public:	11/28/2022 01:07 AM (PT)
Available at:	https://aspredicted.org/tq3m3.pdf

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

We want to explore the relationship between perceptions of Artificial Intelligence and privacy concerns / personal information disclosure.

3) Describe the key dependent variable(s) specifying how they will be measured.

We will assess:

- *Mind perception (7 items)*
- *Technology-specific information disclosure (5-7 items, depending on technology)*
- *Frequency of use (1-3 items).*

For additional measures see below (point 5).

[4] Large parts of this Appendix have been included in the Online Supplement (available at: https://doi.org/10.25384/SAGE.21663916.v1) of the published version of the manuscript constituting Chapter 2 of this dissertation.

4) How many and which conditions will participants be assigned to?

Participants will be randomly assigned to one of five devices (between-subject): voice assistant, search engine, autonomous vehicle, streaming service, washing machine.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Exploratory factor analyses separately for:

- *Action regulation (24 items)*
- *Interactivity (18 items)*
- *Privacy concerns & information disclosure (10 items)*
- *Competence & warmth (12 items)*
- *Trust in provider (5 items).*

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Inclusion requirements for participants:

- *at least 18 years old*
- *survey completion using computer or tablet.*

Participants will be excluded if they fail both attention checks (Attention Check Items: It is important that you pay attention to this study. Please tick "Strongly agree"/ "Strongly disagree"). Participants will also be excluded if they state that they do not consider their own data useable.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We are aiming to sample N = 300 valid cases. To account for data exclusions based on the above described criteria, we will collect data from N = 400 participants (i.e., oversampling of 33%).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

-

Deviations from Pre-Registration

We pre-registered the following analyses which are not reported in the manuscript: separate exploratory factor analyses for (1) action regulation competencies (including all 24 items), (2) interactivity, (3) privacy concerns and willingness to disclose, (4) competence and warmth, (5) trust in provider (see Tables A1-5). We did not include these analyses in the main

text (but below) as they are not primarily relevant to the research question we focused on in the following studies and we thus discuss in the manuscript.

Further, we mistakenly pre-registered to assess five items for trust in provider but actually assessed nine items for this construct. In the manuscript, we refer to ‘conversational AI’ instead of the term ‘voice assistant’ and to ‘willingness to disclose’ instead of ‘information disclosure’ used in the pre-registration.

Table A1

Results From an Initial Factor Analysis for Action Regulation Competencies (Study 1.1: N = 358)

Item	Factor				Scale[a]
	1	2	3	4	
1 ... solves clearly defined tasks in a specific domain.		**.63**			S
2 ... acts according to predefined rules.		**.66**	-.22		S
3 ... considers information that was explicitly submitted for task completion.		**.58**		.25	S
4 ... behaves in a preprogrammed manner.		**.68**	-.17	-.11	S
5 ... executes specific commands.	-.12	**.68**			S
6 ... strictly follows invariant routines. [b]	-.14	**.45**	.33	**-.40**	S
7 ... is prepared to deal with a set of specific situations. [b]		**.48**		.32	F
8 ... adjusts its behavior depending on situational factors.	.36		.35	**.56**	F
9 In different situations, ... behaves differently.		.11	.19	**.67**	F
10 ... uses information that was not explicitly submitted for task completion. [b]	**.66**	-.12		.18	F
11 ... uses contextual information. [b]	**.52**		.23	.25	F
12 ...'s behavior differs depending on the given task.		.35	.16	**.59**	F
13 ... plans its actions.			**.65**	.12	I
14 ... anticipates potential problems.	.33		**.54**	.22	I
15 While completing a task, ... adapts to changing requirements. [b]	**.42**		.16	**.51**	I
16 ... can deal with uncertainty. [b]	.38	-.25	**.43**	.35	I
17 ... makes decisions based on incomplete information. [b]	**.62**	-.12		.13	I
18 ... makes decisions based on probabilistic information. [b]	**.77**		.16	.11	I
19 ... learns from mistakes.	**.65**		.33		M
20 ... adapts its behavior based on prior events.	**.82**		.13		M
21 ... can reason. [b]			**.76**		M
22 ... transfers knowledge to other domains.	**.68**		.26	-.12	M
23 ... solves unknown problems.	.33	-.25	**.55**		M
24 ... finds innovative solutions.	.37		**.60**	.14	M

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings <.1 are not displayed. Factor loadings above .40 are in bold. "…" was replaced with the type of technology in each condition (the voice assistant, the washing machine, the search engine, the streaming service, the autonomous vehicle).
[a] Scale item was designed for: S = sensorimotor level, F = level of flexible action patterns, I = intellectual level, M = level of meta-cognitive heuristics. [b] Item not included in final scales.

Table A2

Results From a Factor Analysis for Interactivity (Study 1.1: N = 358)

Item		Factor		
		1	2	3
1	I can choose freely what ... does.		**0.77**	0.10
2	I have little influence on ...'s behavior. [a]		**0.47**	-0.36
3	I can completely control		**0.77**	
4	While interacting with ..., I have absolutely no control over what happens. [a]	0.22	**0.50**	-0.38
5	My actions determine ...'s behavior.		**0.51**	
6	I have a lot of control over my interaction experience with	0.25	**0.69**	0.11
7	... facilitates two-way communication between itself and its user.	0.20		**0.65**
8	... makes me feel like it wants to interact with me.			**0.77**
9	Using ... is very interactive.	0.22	0.11	**0.60**
10	... is completely apathetic. [a]		-0.27	**0.21**
11	... always reacts to my requests.	**0.50**	**0.43**	
12	Both, ... and I, can start interactions with one another.			**0.74**
13	... processes new input quickly.	0.73		0.12
14	...'s reactions come without delay.	**0.71**	0.13	0.11
15	... responds very slow. [a]	**0.66**		-0.28
16	I never have to wait for ...'s output.	**0.48**		0.38
17	... immediately answers to requests.	**0.71**	0.23	0.14
18	When I interact with ..., I get instantaneous feedback.	**0.65**		0.34

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings <.1 are not displayed. Factor loadings above .40 are in bold. "…" was replaced with the type of technology in each condition (the voice assistant, the washing machine, the search engine, the streaming service, the autonomous vehicle).

[a] Item reverse coded

Table A3

Results From a Factor Analysis for Disclosure (Study 1.1: N = 358)

Item		Factor		Scale[a]
		1	2	
1	I feel that ...'s practices are an invasion of privacy.	-.11	**.86**	PC
2	I feel uncomfortable about the types of information that ... collects.	-.17	**.82**	PC
3	The way that ... monitors its users makes me feel uneasy.	-.16	**.80**	PC
4	I feel personally invaded by the methods used by ... to collect information.	-.18	**.79**	PC
5	I have little reason to be concerned about my privacy when using ... . [R] [b, c]	**-.46**	**.40**	-
6	I would find it acceptable if ... records and uses information about my usage behavior.	**.67**	-.30	ID
7	I would provide ... access to information about me that is stored in or collected by other technological applications or systems.	**.77**	-.20	ID
8	I would provide a lot of information to ... about things that represent me personally.	**.85**		ID
9	I would find it acceptable if ... had a detailed profile of my person.	**.82**	-.19	ID
10	I would give ... access to a lot of information that would characterize me as a person.	**.87**		ID

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings <.1 are not displayed. Factor loadings above .40 are in bold. "…" was replaced with the type of technology in each condition (the voice assistant, the washing machine, the search engine, the streaming service, the autonomous vehicle).
[a] Scale item was included in: PC = Privacy Concerns, ID = Willingness to disclose. [b] Item reverse coded. [c] Item not included in final scales.

Table A4

Results From a Factor Analysis for Competence and Warmth (Study 1.1: N = 358)

Item		Factor	
		1	2
1	... is competent.	0.21	**0.79**
2	... is confident.	**0.72**	0.20
3	... is capable.	0.18	**0.82**
4	... is efficient.		**0.76**
5	... is intelligent.	**0.53**	**0.52**
6	... is skillful.	0.34	**0.63**
7	... is friendly.	**0.81**	0.20
8	... is well-intentioned.	**0.75**	0.26
9	... is trustworthy.	**0.46**	**0.43**
10	... is warm.	**0.75**	
11	... is good-natured.	**0.81**	0.20
12	... is sincere.	**0.78**	0.18

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings <.1 are not displayed. Factor loadings above .40 are in bold. "…" was replaced with the type of technology in each condition (the voice assistant, the washing machine, the search engine, the streaming service, the autonomous vehicle).

Table A5

Results From a Factor Analysis for Trust in Provider (Study 1.1: N = 358)

Item		Factor	
		1	2
1	I believe that ... would act in the best interest of its customers.	**0.85**	0.26
2	... is interested in the well-being of its customers, not just its own.	**0.86**	0.19
3	... is truthful in its dealing with its customers.	**0.82**	0.31
4	I would characterize ... as honest.	**0.86**	0.26
5	... would keep its commitments.	**0.64**	**0.49**
6	... is sincere and genuine.	**0.88**	0.23
7	... is competent.	0.17	**0.85**
8	... performs its role very well.	0.33	**0.76**
9	Overall, ... is capable and proficient.	0.25	**0.86**

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings above .40 are in bold. "…" was replaced with a provider depending on the experimental condition (i.e., Amazon, the provider of the washing machine, Google, Netflix, Tesla).

Additional Analyses

As described in the manuscript, we calculated a second EFA for action regulation after excluding items with ambiguous first loadings, substantial cross loadings and/or loading highest on a factor they were not meant to capture, except for two items which were initially intended for meta-cognitive heuristics but included in the intellectual competencies scale based on the EFA results and reconsideration of their content (see Table A6)

Table A6

Results From a Factor Analysis for the Final Scales of Action Regulation Competencies (Study 1.1: N = 358)

Item	Factor				Scale[a]
	1	2	3	4	
... solves clearly defined tasks in a specific domain.	**.61**		.13	.16	S
... acts according to predefined rules.	**.67**		-.27		S
... considers information that was explicitly submitted for task completion.	**.66**			.12	S
... behaves in a preprogrammed manner.	**.69**		-.25		S
... executes specific commands.	**.71**	-.14			S
... adjusts its behavior depending on situational factors.		.34	.24	**.63**	F
In different situations, ... behaves differently.		.11		**.84**	F
...'s behavior differs depending on the given task.	.32	-.10	.17	**.67**	F
... anticipates potential problems.		.31	**.67**	.20	I
... solves unknown problems.	-.24	.31	**.58**		I
... finds innovative solutions.	-.11	.42	**.56**	.18	I
… can plan actions.			**.79**		I
... learns from mistakes.	-.11	**.76**	.21	.11	M
... adapts its behavior based on prior events.		**.86**			M
... transfers knowledge to other domains.		**.74**	.20		M

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings <.1 are not displayed. Factor loadings above .50 are in bold. "…" was replaced with the type of technology in each condition (the voice assistant, the washing machine, the search engine, the streaming service, the autonomous vehicle).

[a] Scale item was included in: S = sensorimotor level, F = level of flexible action patterns, I = intellectual level, M = level of meta-cognitive heuristics.

Besides, we calculated bivariate correlations between the four levels of action regulation competencies and the two measures of disclosure in the conversational AI condition (see Table A7). Further, we calculated the regression analyses in the conversational AI condition separately for users and non-users (see Table A8) to include Study 1.1 in the meta-analyses across studies.

Table A7

Bivariate Correlations Between Competence Levels and Disclosure (Study 1.1, Conversational AI Condition: n = 72, Non-User Subsample: n = 27, User Subsample: n = 21)

Conversational AI condition (*n* = 72)	**1**	**2**	**3**	**4**	**5**	**6**
1 Sensorimotor level	1.00	0.13	-0.02	0.20	0.09	-0.10
2 Flexible action patterns	0.13	1.00	0.31	0.31	0.20	-0.02
3 Intellectual level	-0.02	0.31	1.00	-0.04	0.49	-0.35
4 Meta-cognitive heuristics	0.20	0.31	-0.04	1.00	-0.19	0.24
5 Willingness to disclose	0.09	0.20	0.49	-0.19	1.00	-0.64
6 Privacy concerns	-0.10	-0.02	-0.35	0.24	-0.64	1.00
Non-users (*n* = 27)	**1**	**2**	**3**	**4**	**5**	**6**
1 Sensorimotor level	1.00	0.11	-0.31	0.20	0.00	-0.06
2 Flexible action patterns	0.11	1.00	0.31	0.53	0.12	0.04
3 Intellectual level	-0.31	0.31	1.00	-0.07	0.28	-0.29
4 Meta-cognitive heuristics	0.20	0.53	-0.07	1.00	-0.21	0.14
5 Willingness to disclose	0.00	0.12	0.28	-0.21	1.00	-0.82
6 Privacy concerns	-0.06	0.04	-0.29	0.14	-0.82	1.00
Users (*n* = 21)	**1**	**2**	**3**	**4**	**5**	**6**
1 Sensorimotor level	1.00	-0.06	0.09	0.09	0.22	-0.17
2 Flexible action patterns	-0.06	1.00	0.46	0.64	0.09	0.33
3 Intellectual level	0.09	0.46	1.00	0.47	0.47	0.18
4 Meta-cognitive heuristics	0.09	0.64	0.47	1.00	0.16	0.36
5 Willingness to disclose	0.22	0.09	0.47	0.16	1.00	-0.38
6 Privacy concerns	-0.17	0.33	0.18	0.36	-0.38	1.00

Table A8

Regressions of Disclosure on Competence Levels (Study 1.1, Conversational AI Condition, Non-User Subsample: n = 27, User Subsample: n = 21)

Sample	Criterion	Predictor	β	*T*	*p*	95%-CI
Non-users (*n* = 27)	Willingness to disclose	Sensorimotor level	0.11	0.54	.594	[-0.33, 0.56]
		Flexible action patterns	0.22	0.85	.406	[-0.31, 0.74]
		Intellectual level	0.22	0.98	.337	[-0.25, 0.70]
		Meta-cognitive heuristics	-0.33	-1.36	.189	[-0.83, 0.17]
	Privacy concerns	Sensorimotor level	-0.22	-1.01	.325	[-0.66, 0.23]
		Flexible action patterns	0.14	0.53	.599	[-0.40, 0.67]
		Intellectual level	-0.39	-1.69	.105	[-0.87, 0.09]
		Meta-cognitive heuristics	0.08	0.35	.731	[-0.42, 0.59]
Users (*n* = 21)	Willingness to disclose	Sensorimotor level	0.17	0.77	.453	[-0.30, 0.63]
		Flexible action patterns	-0.13	-0.44	.665	[-0.74, 0.49]
		Intellectual level	0.52	2.06	.056	[-0.02, 1.05]
		Meta-cognitive heuristics	-0.01	-0.05	.961	[-0.63, 0.60]
	Privacy concerns	Sensorimotor level	-0.18	-0.80	.438	[-0.67, 0.31]
		Flexible action patterns	0.14	0.45	.661	[-0.51, 0.79]
		Intellectual level	0.00	0.00	.997	[-0.56, 0.56]
		Meta-cognitive heuristics	0.29	0.95	.357	[-0.36, 0.94]

Materials of Study 1.1

Procedure (core measures in bold):

1. Informed consent
2. Introductory text for technology (see below)
3. **Perceived competencies** (24 items, displayed in Table A1)
4. Perceived interactivity (18 items), perceived mind (7 items), attention check (1 item)
5. Competence and warmth (12 items)
6. **Privacy concerns and willingness to disclose** (10 items, displayed in Table A3), technology-specific information disclosure (5-7 items, depending on technology), attention check (1 item)
7. Trust in provider (9 items)
8. Usage of respective technology (1-3 items, depending on technology)
9. Demographics, final consent, option for comment

Introductory texts for technology:

a) Voice assistant

We would like you to share your perceptions of voice assistants with us. By voice assistants, we mean technologies that can communicate with their users via voice. One popular example for a voice assistant is Amazon Alexa. Voice assistants are often used via smart speakers which are usually placed in the home of their users. An interaction with a voice assistant can be started by using a wake word such as "Alexa". Afterwards, the voice assistant can react to user requests, such as

- *assessing information from the internet,*
- *streaming different types of media,*
- *setting alarms,*
- *shop online,*
- *controlling other smart devices (e.g., smart light bulbs).*

In addition, skills (or third-party apps) can be installed to enlarge the capabilities of the voice assistant – for example, skills are available that enable the voice assistant to tell jokes or to give compliments to its users. The more the voice assistant is used, the more the system learns about its users' preferences and speech habits. No matter how familiar you are with such a technology or how often you use it – we are interested in your perceptions of voice assistants.

b) Search engine

We would like you to share your perceptions of search engines with us. By search engines, we mean technologies that are used to search the internet for information. One popular example for a search engine is Google. Search engines are often used via computers or smartphones. Typically, users enter one or more key word(s) into a text field to communicate their search query to the search engine. While the user is typing, the search engine makes proposals for potential search queries. After the user submitted a search query, the search engine systematically searches the Internet for matching information. After split seconds, the search engine displays a list of results on the screen. These results may include, for example, web pages, images, videos, or news. We are interested in your perceptions of search engines.

c) Autonomous vehicle

We would like you to share your perceptions of autonomous (or self-driving) cars with us. By autonomous cars, we mean cars that are capable of driving without human

input. A car manufacturer that is well-known for its proceedings in the field of autonomous driving is Tesla. A fully autonomous car completes the driving task without input from a human driver. This driving task includes, amongst others, speed regulation, steering, parking, and navigation. Therefore, the human's role changes from a driver operating the car to a passenger who can engage in different tasks while being chauffeured. This means that the car is also capable of driving when no human is on board.

An autonomous car uses several sensors including cameras, radar and ultrasonic sensors. With these sensors, the autonomous car can, for example, determine the positions of other traffic participants, obstacles and itself. No matter how familiar you are with such a technology or how often you use it – we are interested in your perceptions of autonomous vehicles.

d) Streaming service

We would like you to share your perceptions of streaming services with us. By streaming services, we mean technologies that offer their subscribers a large number of videos on demand. Movies, television programs and series can be streamed online using different devices, such as Smart TVs, smartphones, tablets or computers. A well-known streaming service is Netflix. Using the streaming service, one can either search for a specific title or browse through several proposed categories of movies and series. Streaming services also makes suggestions to their users based on their previously watched videos. The more a streaming service is used, the more it learns about its user's watching preferences and habits. No matter how familiar you are with such a technology or how often you use it – we are interested in your perceptions of streaming services.

e) Washing machine

We would like you to share your perceptions of washing machines with us. By washing machines, we mean home appliances that are used to clean laundry. Usually, the user puts dirty laundry into the washing machine and adds washing soda. Then, the user can choose between washing programs with different water temperatures, washing duration, energy consumption, etc.

Some newer washing machines have sensors to measure how dirty the washing water is and adapt the remaining washing duration based on this information. When the washing process is finished, the user can get the clean laundry out of the washing

machine. No matter how familiar you are with such a technology or how often you use it – we are interested in your perceptions of modern washing machines.

Additional Information for Study 1.2

Exclusion of Participants

In total, the survey was started by 527 participants, of which 425 participants matched the pre-screening (i.e., owning home assistant / smart hub and having used a conversational AI before), completed the survey, and gave consent to use their data. 22 participants were excluded from analysis as they participated in the survey more than once. Based on our pre-registration (https://aspredicted.org/TP3_1XF) we further excluded participants who (1) did not complete the survey using a computer or tablet, (2) did not pass both attention checks, and/or (3) did not consider their own data useable, which led to the exclusion of 68 participants. The pre-registered exclusion criterion of an $SDR > 2.59$ in the regression analyses testing the main prediction lead to one additional exclusion. Thus, the eligible sample consisted of $N = 334$ participants (as reported in the manuscript).

Pre-Registration

Created:	05/21/2021 12:20 AM (PT)
Made public:	11/28/2022 01:12 AM (PT)
Available at:	https://aspredicted.org/v87ub.pdf

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

H1: Higher perceptions of warmth in a voice assistant predict (a) a higher willingness to disclose information and (b) lower privacy concerns.

H2: Higher perceived cognitive competencies on the intellectual level in a voice assistant predict (a) a higher willingness to disclose information and (b) lower privacy concerns.

H3: Higher perceived competencies on the level of meta-cognitive heuristics in a voice assistant predict (a) a lower willingness to disclose information and (b) higher privacy concerns.

H4: Higher perceptions of reciprocal interaction between the user and the voice assistant predict (a) a higher willingness to disclose information and (b) lower privacy concerns.

RQ1: Do lower order cognitive competencies (sensorimotor, flexible action patterns) in a voice assistant affect (a) willingness to disclose and (b) privacy concerns?

3) Describe the key dependent variable(s) specifying how they will be measured.

- *Privacy concerns (5 items),*
- *Willingness to disclose information (5 items).*

4) How many and which conditions will participants be assigned to?

None. We will assess the predictors as specified under #5.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will conduct separate multiple regression analyses regressing (a) willingness to disclose information and (b) privacy concerns separately on

(1) Four dimensions of cognitive competencies,

(2) Competence and warmth,

(3) Three dimensions of interactivity.

We will check whether the observed effects hold when controlling for trust in the system provider, participant age and gender by including these three variables in each of the regressions described before.

In advance of the regression analyses, we will conduct exploratory factor analyses with varimax rotations since the used scales are not yet well-established. Scales will be formed based on highest item loadings on the predicted factors. We will calculate factor analyses separately for:

- *Cognitive competencies: rule-based task completion (6 items), flexible action patterns (6 items), intellect (6 items), meta-cognitive heuristics (6 items);*
- *Interactivity: active control (6 items), two-way interaction (6 items), synchronicity (6 items);*
- *Competence (4 items) & warmth (5 items);*
- *Privacy concerns (5 items) & willingness to disclose information (5 items);*
- *Trust in system provider: benevolence/integrity (6 items), competence (3 items).*

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Inclusion requirements for participants:

- *at least 18 years old,*
- *consent for participation and usage of data for scientific purposes,*
- *survey completion using computer or tablet,*

- *passing two attention checks (Attention Check Items: It is important that you pay attention to this study. Please tick "Strongly agree"/ "Strongly disagree"),*
- *experience with the usage of a voice assistant.*

Participants will be excluded if they state that they do not consider their own data useable. After excluding participants according to these criteria, data will be checked for outliers using studentized deleted residuals (SDR) from the multiple regression analysis described under #5a1. Participants with an absolute SDR > 2.59 will be regarded as statistical outliers.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We seek for a power of 95% for observing effects with a minimum effect size of $f^2 = .03$ in a multiple regression analysis with four predictors, hence, we aim to collect N = 340 valid cases. To account for data exclusions based on the above described criteria, we will collect data from N = 425 participants (i.e., oversampling of 25%).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will exploratory check for interactions between warmth and the levels of cognitive competencies. We refrain from formulating hypotheses for these interactions as the size of correlation between the interacting factors has not yet been determined.

Further, we will exploratory assess:

- *Agentic mind perception (7 items),*
- *Technology-specific information disclosure (6 items),*
- *Usefulness (4 items), ease of use (4 items), enjoyment (3 items),*
- *Frequency and purpose of use (1-6 items).*

Deviations from Pre-Registration

We pre-registered two additional hypotheses (H1 and H4 in the pre-registration) regarding the effects of warmth and interactivity on disclosure, which we did not include in the manuscript as they are not central for the research question as it evolved over the course of the studies and discussed in the main text.

Besides, we pre-registered that the first level of cognitive competencies would be referred to as ‘rule-based task completion’. Within the research process, we adjusted this label to be in agreement with the original label from the action regulation theory and, thus, refer to it as

the ‘sensorimotor level’ in the paper. Further, in the manuscript we refer to ‘conversational AI’ instead of the term ‘voice assistant’ used in the pre-registration.

We further pre-registered several additional analyses. First, exploratory factor analyses with varimax rotations for (1) cognitive competencies (i.e., action regulation competencies), (2) interactivity, (3) competence and warmth, (4) privacy concerns and willingness to disclose, (5) trust in provider. Second, multiple regression analyses regressing (a) willingness to disclose and (b) privacy concerns separately on (1) competence and warmth and (2) three dimensions of interactivity. Third, we pre-registered to check whether the effects observed in the regression analyses hold when controlling for trust in the system provider, participant age and gender by including these three variables in the respective regressions. Fourth, we pre-registered to exploratory check for interactions between warmth and the competence levels derived from the action regulation theory. We did not include these analyses in the paper as they were not central for the discussed research questions and/or did not yield clear patterns of results. Instead, these analyses are reported below (see Table A9 to Table A17).

Table A9

Results From a Factor Analysis for Action Regulation Competencies (Study 1.2: N = 334)

Item		Factor				Scale[a]
		1	2	3	4	
1	The voice assistant solves clearly defined problems in a specific domain.		.27	**.65**		S
2	The voice assistant acts according to predefined rules.			**.71**		S
3	The voice assistant considers information that was explicitly submitted for task completion.	.10	.22	**.54**		S
4	The voice assistant behaves in a preprogrammed manner.			**.58**	-.18	S
5	The voice assistant executes specific commands.			**.64**		S
6	The voice assistant is prepared to deal with specific situations.	.12		**.57**	.31	S
7	The voice assistant adjusts its behavior depending on situational factors.	**.61**	.27		.27	F
8	In different situations, the voice assistant behaves differently.	**.77**	.14			F
9	The voice assistant's behavior differs depending on the given task.	**.68**		.18	.12	F
10	The voice assistant's behavior depends on the given context.	**.67**	.17			F
11	The voice assistant adapts its behavior depending on who is interacting with it.	**.62**	.33	-.14	.12	F
12	The voice assistant's behavior is sensitive to environmental factors.	**.57**			.34	F
13	The voice assistant anticipates potential problems.	.17	.34	-.14	**.52**	I
14	The voice assistant can deal with incomplete information.	.12			**.69**	I
15	The voice assistant solves unknown tasks.		.37	-.21	**.44**	I
16	The voice assistant finds innovative solutions.	.15	**.43**	-.16	**.51**	I
17	The voice assistant can plan actions.	.22	.28	.13	.35	I
18	The voice assistant can extract key information from complex input.	.21		.27	**.66**	I
19	The voice assistant learns from mistakes.	.28	**.68**	-.14	.22	M
20	The voice assistant adapts its behavior based on prior events.	.34	**.73**		.16	M
21	The voice assistant transfers knowledge to other domains.		**.44**	.25		M
22	Based on previous incidents, the voice assistant derives universal strategies.		**.46**		.38	M
23	The voice assistant develops general strategies to accomplish various tasks.		**.40**		**.46**	M
24	The voice assistant uses information from prior occurrences to solve related tasks.	.18	**.73**	.18	.20	M

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings <.1 are not displayed. Factor loadings above .40 are in bold.

[a] Scale item was designed for: S = sensorimotor level, F = level of flexible action patterns, I = intellectual level, M = level of meta-cognitive heuristics

Table A10

Results From a Factor Analysis for Interactivity (Study 1.2: N = 334)

Item		Factor		
		1	2	3
1	I can choose freely what the voice assistant does.	0.28	0.27	**0.57**
2	I have little influence on the voice assistant's behavior. [a]		-0.10	**0.73**
3	I can completely control the voice assistant.	0.26	0.18	**0.56**
4	While interacting with the voice assistant, I have absolutely no control over what happens. [a]		-0.14	**0.68**
5	My actions determine the voice assistant's behavior.		0.38	0.32
6	I have a lot of control over my interaction experience with the voice assistant.	0.28	0.33	**0.62**
7	The voice assistant facilitates two-way communication between itself and its user.	0.10	**0.67**	0.24
8	The voice assistant makes me feel like it wants to interact with me.	0.16	**0.69**	
9	Using the voice assistant is very interactive.	**0.40**	**0.55**	0.21
10	The voice assistant gives me the opportunity to interact with it.	0.14	**0.55**	0.31
11	The voice assistant encourages me to interact with it	0.11	**0.68**	
12	Both, the voice assistant and I, can start interactions with one another.		**0.66**	-0.10
13	The voice assistant processes new input quickly.	**0.57**	0.30	0.25
14	The voice assistant's reactions come without delay.	**0.77**	0.18	
15	The voice assistant responds very slow. [a]	**0.69**		0.15
16	I never have to wait for the voice assistant's output.	**0.68**	0.12	
17	The voice assistant immediately answers to requests.	**0.79**	0.11	0.14
18	When I interact with the voice assistant, I get instantaneous feedback.	**0.77**	0.20	0.10

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings <.1 are not displayed. Factor loadings above .40 are in bold.

[a] Item reverse coded

Table A11

Results From a Factor Analysis for Competence and Warmth (Study 1.2: N = 334)

Item		Factor	
		1	2
1	The voice assistant is competent.	0.28	**0.82**
2	The voice assistant is capable.	0.19	**0.83**
3	The voice assistant is efficient.		**0.81**
4	The voice assistant is skillful.	0.28	**0.77**
5	The voice assistant is friendly.	**0.71**	0.31
6	The voice assistant is well-intentioned.	**0.76**	0.28
7	The voice assistant is warm.	**0.77**	0.10
8	The voice assistant is good-natured.	**0.82**	0.16
9	The voice assistant is sincere.	**0.76**	0.16

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings <.1 are not displayed. Factor loadings above .40 are in bold.

Table A12

Results From a Factor Analysis for Privacy Concerns and Willingness to Disclose (Study 1.2: N = 334)

Item		Factor		Scale[a]
		1	2	
1	I feel that the voice assistant's practices are an invasion of privacy.	**0.84**	-0.24	PC
2	I feel uncomfortable about the types of information that the voice assistant collects.	**0.79**	-0.26	PC
3	The way that the voice assistant monitors its users makes me feel uneasy.	**0.84**	-0.21	PC
4	I feel personally invaded by the methods used by the voice assistant to collect information.	**0.84**	-0.17	PC
5	I am concerned about my privacy when using the voice assistant	**0.79**	-0.21	PC
6	I would find it acceptable if the voice assistant records and uses information about my usage behavior.	-0.22	**0.77**	ID
7	I would provide the voice assistant access to information about me that is stored in or collected by other technological applications or systems.	-0.16	**0.80**	ID
8	I would provide a lot of information to the voice assistant about things that represent me personally.	-0.19	**0.83**	ID
9	I would find it acceptable if the voice assistant had a detailed profile of my person.	-0.30	**0.79**	ID
10	I would give the voice assistant access to a lot of information that would characterize me as a person.	-0.23	**0.87**	ID

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings <.1 are not displayed. Factor loadings above .40 are in bold.

[a] Scale item was included in: PC = Privacy Concerns, ID = Willingness to disclose

Table A13

Results From a Factor Analysis for Trust in Provider (Study 1.2: N = 334)

Item		Factor	
		1	2
1	I believe that the provider would act in the best interest of its customers.	**0.86**	0.21
2	The provider is interested in the well-being of its customers, not just its own.	**0.86**	0.21
3	The provider is truthful in its dealing with its customers.	**0.82**	0.25
4	I would characterize the provider as honest.	**0.85**	0.28
5	The provider would keep its commitments.	**0.58**	**0.51**
6	The provider is sincere and genuine.	**0.84**	0.25
7	The provider is competent.	0.23	**0.84**
8	The provider performs its role very well.	0.21	**0.86**
9	Overall, the provider is capable and proficient.	0.27	**0.86**

Note. The extraction method was principal axis factoring with an orthogonal (varimax) rotation and Kaiser normalization. Factor loadings above .40 are in bold.

Table A14

Regressions of Disclosure on Competence Levels and Control Variables (Study 1.2: N = 334)

Predictor	Willingness to disclose			Privacy concerns		
	β	*p*	95%-CI	β	*p*	95%-CI
Sensorimotor level	**0.23**	<.001	[0.12, 0.34]	**-0.16**	.006	[-0.27, -0.05]
Flexible action patterns	0.09	.130	[-0.03, 0.21]	0.04	.557	[-0.08, 0.16]
Intellectual level	0.12	.084	[-0.02, 0.25]	-0.03	.630	[-0.17, 0.10]
Meta-cognitive heuristics	-0.07	.290	[-0.20, 0.06]	**0.19**	.007	[0.05, 0.32]
Provider benevolence / integrity	**0.34**	<.001	[0.21, 0.46]	**-0.32**	<.001	[-0.44, -0.19]
Provider competence	-0.06	.361	[-0.19, 0.07]	-0.07	.308	[-0.20, 0.06]
Sensorimotor level	**0.21**	<.001	[0.11, 0.31]	**-0.19**	.001	[-0.30, -0.08]
Flexible action patterns	0.08	.184	[-0.04, 0.20]	0.06	.382	[-0.07, 0.18]
Intellectual level	**0.20**	.002	[0.07, 0.33]	**-0.14**	.048	[-0.28, -0.001]
Meta-cognitive heuristics	-0.05	.485	[-0.18, 0.09]	**0.15**	.036	[0.01, 0.29]
Participant age	**-0.19**	<.001	[-0.29, -0.09]	0.05	.322	[-0.05, 0.16]
Sensorimotor level	**020**	<.001	[0.09, 0.31]	**-0.17**	.002	[-0.28, -0.06]
Flexible action patterns	0.06	.346	[-0.06, 0.18]	0.06	.331	[-0.06, 0.19]
Intellectual level	**0.23**	.001	[0.10, 0.37]	**-0.16**	.024	[-0.30, -0.02]
Meta-cognitive heuristics	-0.04	.552	[-0.18, 0.10]	**0.15**	.038	[0.01, 0.29]
Participant gender	0.08	.135	[-0.3 0.19]	0.03	.637	[-0.08, 0.14]

Note. Significant coefficients in bold.

Table A15

Regressions of Disclosure on Competence, Warmth, and Control Variables (Study 1.2: N = 334)

	Willingness to disclose			Privacy concerns		
Predictor	β	*p*	95%-CI	β	*p*	95%-CI
Competence	**0.34**	<.001	[0.22, 0.45]	**-0.19**	.002	[-0.31, -0.07]
Warmth	0.08	.193	[-0.04, 0.19]	-0.09	.137	[-0.21, 0.03]
Competence	**0.26**	<.001	[0.14, 0.38]	-0.10	.099	[-0.23, 0.02]
Warmth	0.01	.859	[-0.11, 0.14]	0.03	.677	[-0.10, 0.16]
Provider benevolence / integrity	**0.23**	<.001	[0.11, 0.35]	**-0.22**	<.001	[-0.35, -0.09]
Provider competence	0.00	.961	[-0.12, 0.13]	-0.13	.055	[-0.26, 0.003]
Competence	**0.35**	<.001	[0.24, 0.46]	**-0.19**	.002	[-0.31, -0.07]
Warmth	0.07	.191	[-0.04, 0.19]	-0.09	.139	[-0.21, 0.03]
Participant age	**-0.23**	<.001	[-0.33, -0.13]	0.07	.187	[-0.03, 0.18]
Competence	**0.33**	<.001	[0.21, 0.44]	**-0.18**	.004	[-0.30, -0.06]
Warmth	0.10	.092	[-0.02, 0.21]	-0.11	.064	[-0.24, 0.01]
Participant gender	**0.10**	.047	[0.00, 0.20]	0.03	.625	[-0.08, 0.13]

Note. Significant coefficients in bold.

Table A16

Regressions of Disclosure on Interactivity and Control Variables (Study 1.2: N = 334)

	Willingness to disclose			Privacy concerns		
Predictor	β	*p*	95%-CI	β	*p*	95%-CI
Active control	0.03	.654	[-0.09, 0.14]	-0.07	.248	[-0.19, 0.05]
Two-way interaction	**0.27**	<.001	[0.16, 0.38]	-0.07	.224	[-0.19, 0.04]
Synchronicity	0.03	.571	[-0.08, 0.15]	-0.11	.073	[-0.23, 0.01]
Active control	0.00	.980	[-0.11, 0.11]	-0.04	.525	[-0.15, 0.08]
Two-way interaction	**0.20**	.001	[0.08, 0.31]	0.01	.848	[-0.10, 0.12]
Synchronicity	-0.01	.813	[-0.13, 0.10]	-0.04	.509	[-0.16, 0.08]
Provider benevolence / integrity	**0.29**	<.001	[0.17, 0.41]	**-0.25**	<.001	[-0.37, -0.12]
Provider competence	0.01	847	[-0.11, 0.14]	-0.11	.089	[-0.24, 0.02]
Active control	0.03	.654	[-0.09, 0.14]	-0.07	.249	[-0.19, 0.05]
Two-way interaction	**0.28**	<.001	[0.17, 0.38]	-0.07	.217	[-0.19, 0.04]
Synchronicity	0.06	.283	[-0.05, 0.18]	-0.12	.051	[-0.24, 0.0005]
Participant age	**-0.24**	<.001	[-0.34, -0.14]	0.09	.119	[-0.02, 0.19]
Active control	0.02	.717	[-0.09, 0.14]	-0.08	.159	[-0.20, 0.03]
Two-way interaction	**0.27**	<.001	[0.16, 0.39]	-0.07	.216	[-0.19, 0.04]
Synchronicity	0.04	.528	[-0.08, 0.16]	-0.11	.082	[-0.23, 0.01]
Participant gender	0.09	.093	[-0.02, 0.20]	0.03	.535	[-0.07, 0.14]

Note. Significant coefficients in bold.

Table A17

Regressions of Disclosure on Action Regulation Competencies and Warmth (Study 1.2: N = 334)

	Willingness to disclose			Privacy concerns		
Predictor	β	*p*	95%-CI	β	*p*	95%-CI
Sensorimotor level	**0.18**	.001	[0.08, 0.29]	**-0.17**	.002	[-0.28, -0.06]
Flexible action patterns	0.07	.217	[-0.04, 0.19]	0.07	.244	[-0.05, 0.20]
Intellectual level	0.12	.084	[-0.02, 0.25]	-0.04	.532	[0.001, 0.28]
Meta-cognitive heuristics	-0.05	.487	[-0.18, 0.08]	**0.14**	.048	[-0.32, -0.08]
Warmth	**0.30**	<.001	[0.19, 0.41]	**-0.20**	.001	[-0.32, -0.08]
Sensorimotor level x Warmth	0.08	.166	[-0.03, 0.15]	-0.10	.109	[-0.17, 0.02]
Flexible action patterns x Warmth	0.07	.231	[-0.04, 0.18]	0.00	.996	[-0.12, 0.12]
Intellectual level x Warmth	-0.04	.31	[-0.15, 0.09]	**0.17**	.027	[0.02, 0.26]
Meta-cognitive heuristics x Warmth	0.02	.801	[-0.11, 0.14]	-0.01	.884	[-0.14, 0.12]

Note. Significant coefficients in bold.

Additional Analyses

We additionally report bivariate correlations between competence levels and disclosure (see Table A18).

Table A18

Bivariate Correlations Between Competence Levels and Disclosure (Study 1.2: N = 334)

	1	**2**	**3**	**4**	**5**	**6**
1 Sensorimotor level	1.00	0.11	0.11	0.23	0.24	-0.16
2 Flexible action patterns	0.11	1.00	0.48	0.49	0.19	0.04
3 Intellectual level	0.11	0.48	1.00	0.60	0.25	-0.05
4 Meta-cognitive heuristics	0.23	0.49	0.60	1.00	0.17	0.05
5 Willingness to disclose	0.24	0.19	0.25	0.17	1.00	-0.50
6 Privacy concerns	-0.16	0.04	-0.05	0.05	-0.50	1.00

Materials of Study 1.2

Procedure (core measures in bold):

1. Informed consent
2. Prescreening
3. Introductory text (see below)
4. **Perceived competencies** (24 items, displayed in Table 2)
5. Perceived interactivity (18 items), perceived agency (7 items), attention check (1 item)
6. Competence and warmth (9 items)

7. **Privacy concerns and willingness to disclose** (10 items, displayed in Table A12), technology-specific information disclosure (6 items), attention check (1 item)
8. Perceived usefulness, ease of use, and enjoyment (11 items)
9. Frequency and purpose of use (1-6 items)
10. Trust in provider (9 items)
11. Demographics, final consent, option for comment

Introductory text:

> *We would like you to share your perceptions of voice assistants with us. By voice assistants, we mean technologies that can communicate with their users via voice. They can process voice input from their users and produce voice output themselves. Popular examples of voice assistants are Amazon Alexa, Google Assistant or Siri.*
> *After being activated, a typical voice assistant can react to different types of user requests, for example, assessing information from the internet, setting alarms or streaming media. We are interested in your individual perceptions of such voice assistants.*

Additional Information for Study 1.3

Exclusion of Participants

We used Prolific pre-screeners to include only participants who were native English speakers and who had an approval rate of at least 97%. In total, the survey was started by 409 participants of which 401 completed the survey with consent to use their data. As pre-registered (https://aspredicted.org/92W_WVQ), we excluded a total of 77 participants of whom (1) five did not complete the survey using a computer or tablet, (2) 20 did not pass both attention checks, (3) five did not consider their own data useable, and (4) an additional 26 were not clearly classifiable as either a non-user or a user of conversational AI (as defined above). One additional participant was omitted from the data set because they participated twice. The pre-registered exclusion criterion of an $SDR > 2.59$ in the regression analyses testing the main prediction did not lead to any additional exclusions. Thus, the eligible sample consisted of $N = 323$ participants (as reported in the manuscript).

Pre-Registration

Created:	12/02/2021 01:28 AM (PT)
Made public:	11/28/2022 01:15 AM (PT)
Available at:	https://aspredicted.org/qg5m4.pdf

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

H1: Higher competencies on the level of meta-cognitive heuristics lead to (a) lower willingness to disclose and (b) higher privacy concerns, but more so among CAI non-users than among users.

H2: Higher competencies on the intellectual level lead to (a) higher willingness to disclose and (b) lower privacy concerns.

In line with the logic of H1 we would assume that the effect stated in H2 will be stronger among CAI non-users than among users. However, given that our previous findings do not allow to conclude a moderation by usage experience, we do not pre-register this as a hypothesis, but rather plan to explore the interaction pattern.

3) Describe the key dependent variable(s) specifying how they will be measured.

- *Privacy concerns (5 items)*
- *Willingness to disclose information (5 items)*
- *Usage experience (1 item)*

4) How many and which conditions will participants be assigned to?

We will implement a 2 (intellectual competencies: low vs high) x 2 (meta-cognitive competencies: low vs high) x 2 (usage experience: users vs non-users) between-subjects design. Usage experience will be a quasi-experimental factor: Participants will be regarded as users if they indicate to use a voice assistant at least several times a week. Participants will be regarded as non-users if they indicate that they use a voice assistant maximum once a month. Intellectual and meta-cognitive competencies will be experimentally manipulated: participants will be randomly assigned to one of the four resulting experimental conditions implemented via short descriptions of an update for a CAI.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will conduct a MANOVA with the two experimental factors (intellectual competencies & meta-cognitive competencies) and the quasi-experimental factor (usage experience) as independent variables and privacy concerns as well as willingness to disclose information as dependent variables. Interaction effects of competencies (intellectual level / level of meta-cognitive heuristics) by usage experience will be used to

test Hypothesis 1 and the addition to Hypothesis 2. Main effects of competencies on the intellectual level will provide evidence for Hypothesis 2.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Inclusion requirements for participants:

- *at least 18 years old,*
- *consent for participation and usage of data for scientific purposes,*
- *survey completion using computer or tablet,*
- *passing two attention checks,*
- *being either a non-user or a user of CAIs (as specified under point 4 and in line with answers from prescreening).*

Participants will be excluded if they state that they did not answer all questions honestly and attentively. After excluding participants according to these criteria, data will be checked for outliers using studentized deleted residuals (SDR) from a multiple regression analysis of willingness to disclose information on the three factors (usage experience, competencies on the intellectual level, competencies on the level of meta-cognitive heuristics), and their interactions. Participants with an absolute SDR > 2.59 will be regarded as statistical outliers.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We seek for a power of 80% and an alpha error of 0.05 for observing a minimum effect of f2 = 0.02 in a MANOVA with eight groups, three independent and two dependent variables. Hence, we aim to collect N = 344 valid cases. To account for data exclusion based on the above described criteria, we will collect data from N = 400 participants (i.e., oversampling of 15%).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will assess perceived competencies on the intellectual level (3 items) and on the level of meta-cognitive heuristics (3 items) as a manipulation check. We will further exploratory asses perceived warmth of the voice assistant (5 items), general attitude towards CAIs (3 items), trust in the CAI provider (10 items), and whether participants used CAIs more frequently at an earlier point in time (1 item).

Deviations from Pre-Registration

We pre-registered an interaction hypothesis assuming that higher competencies on the level of meta-cognitive heuristics lead to (a) lower willingness to disclose and (b) higher privacy concerns, but more so among conversational AI non-users than among users (H1). This effect was however non-significant (as reported in the manuscript). We did not include this hypothesis in the paper but followed its idea by differentiating between users and non-users within the analyses reported in the paper. Further, we replaced the term 'voice assistant' and the abbreviation 'CAI' used in the pre-registration with the wording 'conversational AI'.

Additional Analyses

We additionally report bivariate correlations between competence levels and disclosure (see Table A19).

Table A19

Bivariate Correlations Between Competence Levels and Disclosure (Study 1.3: N_{total} = 323, $n_{Non\text{-}Users}$ = 167, n_{Users} = 156)

Complete sample (N = 323)	**1**	**2**	**3**	**4**
1 Intellectual level (MC)	1.00	0.51	0.08	0.01
2 Meta-cognitive heuristics (MC)	0.51	1.00	0.00	0.16
3 Willingness to disclose	0.08	0.00	1.00	-0.73
4 Privacy concerns	0.01	0.16	-0.73	1.00
Non-users (n = 167)	**1**	**2**	**3**	**4**
1 Intellectual level (MC)	1.00	0.44	0.01	0.06
2 Meta-cognitive heuristics (MC)	0.44	1.00	-0.10	0.24
3 Willingness to disclose	0.01	-0.10	1.00	-0.73
4 Privacy concerns	0.06	0.24	-0.73	1.00
Users (n = 156)	**1**	**2**	**3**	**4**
1 Intellectual level (MC)	1.00	0.57	0.16	-0.03
2 Meta-cognitive heuristics (MC)	0.57	1.00	0.10	0.10
3 Willingness to disclose	0.16	0.10	1.00	-0.69
4 Privacy concerns	-0.03	0.10	-0.69	1.00

Note. MC = Manipulation check

Materials of Study 1.3

Procedure (core measures in bold):

1. Informed consent
2. Introductory text (see below)
3. Scenario description (see below)

4. Manipulation of intellectual competencies and meta-cognitive heuristics (see below)
5. **Privacy concerns and willingness to disclose** (10 items, same as in Study 1.2), attention check (1 item)
6. **Manipulation check** (6 items, see below), attention check (1 item)
7. Perceived warmth (5 items)
8. Conversational AI usage (1-5 items)
9. Trust in provider (10 items)
10. General attitude towards conversational AI (3 items)
11. Demographics, debriefing, final consent, option for comment

Introductory text:

> *In this study, we are interested in your perceptions of voice assistants. By voice assistants, we mean technologies that can communicate with their users via voice. Such technologies are often used via smart speakers which are usually placed in the home of their users. Popular examples of voice assistants are Amazon Alexa, Google Assistant, and Apple's Siri.*
>
> *An interaction with a typical voice assistant can be started by using a wake word such as "Alexa" or "Hey Google". Afterward, the voice assistant can react to different types of user requests, such as assessing information from the internet, streaming different types of media, setting alarms, shopping online, or controlling other smart devices (e.g., smart light bulbs).*

Scenario description:

> *Please imagine the following situation. In case you do not use a voice assistant, please imagine you would use one.*
>
> *The provider of your voice assistant has announced the development of a major update. In advance of the release, we would like to know how people will react to the changes that come with this update.*
>
> *To get a reasonable impression of what you think about the planned update, we will give you a realistic preview of what the voice assistant will be capable of after the update but also of what the voice assistant will not be able to do despite the update. On the following pages, you will find some short information about the voice assistant's capabilities after the planned update.*
>
> *Please read this information carefully as you will be asked several questions about it afterward.*

(page break)

After the update, the voice assistant will have a substantially improved capability to react to more different user requests by executing specific commands. Another improvement is that it will consider characteristics of specific situations more than before (e.g., behave differently depending on where and when the interaction takes place, or who is interacting with it).

Manipulation of intellectual competencies and meta-cognitive heuristics:

Low intellectual competencies, low meta-cognitive heuristics:

Yet, after installing the update, the voice assistant will NOT have improved its capabilities

- *to gather key information out of complex input, nor*
- *to plan actions in advance, nor*
- *to find innovative solutions.*

Consider the following example: you ask the voice assistant to read out an email in which you are invited to an upcoming event (e.g., a birthday party or a work meeting in another city). If you decide to join the event, your voice assistant cannot give you more useful recommendations on how to get there compared to before the update.

You will have to specify which means of transportation you want to choose (e.g., car, train, or plane) and at what time you want to arrive at your destination. The voice assistant will not automatically evaluate for each option whether it is realistic to start the journey on the same day as the event takes place but will make recommendations solely based on your input.

For example, if the event takes place at noon in a rather distant location and you would like to go there by car, it might propose that you start your journey in the middle of the night instead of leaving the day before and suggesting a hotel to stay at. Besides, you will have to explicitly command the voice assistant to add your estimated travel times as well as the event itself to your calendar.

(page break)

In addition, the voice assistant will NOT have improved its capabilities

- *to use information from prior events to solve related tasks, nor*
- *to develop general strategies for task completion, nor*
- *to learn from its previous mistakes.*

Even after you will have used the voice assistant for some time, the system will not be better capable of learning about your general personal preferences and attitudes. Instead, it will only know about preferences and attitudes that you explicitly submitted.
For example, the voice assistant will not be able to give better-personalized recommendations when asked for a holiday destination unless you explicitly share additional information.
When only asked for a holiday destination recommendation, the voice assistant would not consider your previous activities using the voice assistant (e.g., which products you bought or rated favorably, which events you added to your calendar) but give general advice for nice holiday destinations independent of your prior activities.

High intellectual competencies, low meta-cognitive heuristics:

Besides, after installing the update, the voice assistant will have improved its capabilities

- *to gather key information out of complex input, and*
- *to plan actions in advance, and*
- *to find innovative solutions.*

Consider the following example: you ask the voice assistant to read out an email in which you are invited to an upcoming event (e.g., a birthday party or a work meeting in another city). If you decide to join the event, your voice assistant can now give you more useful recommendations on how to get there. It will not only offer you different options (e.g., car, train, or plane) but evaluate for each option whether it is realistic to start the journey on the same day as the event takes place or whether you should better travel there the day before.
For example, if the event takes place at noon in a rather distant location, it might tell you that you can either go by car the day before the event, take a night train, or take a plane on the same day as the event. In case you decide on an option with which you will arrive at your destination a day before the event, the voice assistant will also give you recommendations for suitable hotels. Besides, it will automatically add your estimated travel times as well as the event itself to your calendar.
(page break)
However, the voice assistant will NOT have improved its capabilities

- *to use information from prior events to solve related tasks, nor*
- *to develop general strategies for task completion, nor*
- *to learn from its previous mistakes.*

Even after you will have used the voice assistant for some time, the system will not be better capable of learning about your general personal preferences and attitudes. Instead, it will only know about preferences and attitudes that you explicitly submitted.
For example, the voice assistant will not be able to give better-personalized recommendations when asked for a holiday destination unless you explicitly share additional information.
When only asked for a holiday destination recommendation, the voice assistant would not consider your previous activities using the voice assistant (e.g., which products you bought or rated favorably, which events you added to your calendar) but give general advice for nice holiday destinations independent of your prior activities.

Low intellectual competencies, high meta-cognitive heuristics:

Yet, after installing the update, the voice assistant will NOT have improved its capabilities

- *to gather key information out of complex input, nor*
- *to plan actions in advance, nor*
- *to find innovative solutions.*

Consider the following example: you ask the voice assistant to read out an email in which you are invited to an upcoming event (e.g., a birthday party or a work meeting in another city). If you decide to join the event, your voice assistant cannot give you more useful recommendations on how to get there compared to before the update.
You will have to specify which means of transportation you want to choose (e.g., car, train, or plane) and at what time you want to arrive at your destination. The voice assistant will not automatically evaluate for each option whether it is realistic to start the journey on the same day as the event takes place but will make recommendations solely based on your input.
For example, if the event takes place at noon in a rather distant location and you would like to go there by car, it might propose that you start your journey in the middle of the night instead of leaving the day before and suggesting a hotel to stay at. Besides, you will have to explicitly command the voice assistant to add your estimated travel times as well as the event itself to your calendar.
(page break)
However, the voice assistant will have improved its capabilities

- *to use information from prior events to solve related tasks, and*
- *to develop general strategies for task completion, and*

- *to learn from its previous mistakes.*

After you have used the voice assistant for some time, the system will better learn about your general personal preferences and attitudes instead of knowing only the preferences and attitudes that you explicitly submitted.

For example, the voice assistant will be able to give you better-personalized recommendations when asked for a holiday destination without further need for input. When asked for a recommendation, the voice assistant will now consider your previous activities using the voice assistant.

If you often buy sports clothing or give favorable ratings for such, add sports events to your calendar or track sportive activities, the voice assistant would likely suggest you a holiday destination suitable for different sportive activities. In contrast, if you rather buy books and give favorable ratings for theater performances, the voice assistant would rather suggest a holiday destination that offers a broad range of cultural activities.

High intellectual competencies, high meta-cognitive heuristics:

Besides, after installing the update, the voice assistant will have improved its capabilities

- *to gather key information out of complex input, and*
- *to plan actions in advance, and*
- *to find innovative solutions.*

Consider the following example: you ask the voice assistant to read out an email in which you are invited to an upcoming event (e.g., a birthday party or a work meeting in another city). If you decide to join the event, your voice assistant can now give you more useful recommendations on how to get there. It will not only offer you different options (e.g., car, train, or plane) but evaluate for each option whether it is realistic to start the journey on the same day as the event takes place or whether you should better travel there the day before.

For example, if the event takes place at noon in a rather distant location, it might tell you that you can either go by car the day before the event, take a night train, or take a plane on the same day as the event. In case you decide on an option with which you will arrive at your destination a day before the event, the voice assistant will also give you recommendations for suitable hotels. Besides, it will automatically add your estimated travel times as well as the event itself to your calendar.

(page break)

In addition, the voice assistant will have improved its capabilities

- *to use information from prior events to solve related tasks, and*
- *to develop general strategies for task completion, and*
- *to learn from its previous mistakes.*

After you have used the voice assistant for some time, the system will better learn about your general personal preferences and attitudes instead of knowing only the preferences and attitudes that you explicitly submitted.

For example, the voice assistant will be able to give you better-personalized recommendations when asked for a holiday destination without further need for input. When asked for a recommendation, the voice assistant will now consider your previous activities using the voice assistant.

If you often buy sports clothing or give favorable ratings for such, add sports events to your calendar or track sportive activities, the voice assistant would likely suggest you a holiday destination suitable for different sportive activities. In contrast, if you rather buy books and give favorable ratings for theater performances, the voice assistant would rather suggest a holiday destination that offers a broad range of cultural activities.

Manipulation check:

Intellectual competencies:

- *The voice assistant anticipates potential problems.*
- *The voice assistant can deal with incomplete information.*
- *The voice assistant solves unknown tasks*

Meta-cognitive heuristics:

- *The voice assistant adapts its behavior based on prior events.*
- *The voice assistant transfers knowledge to other domains.*
- *Based on previous incidents, the voice assistant derives universal strategies.*

Additional Information for Study 1.4

Exclusion of Participants

Participation requirements included not being a regular user of conversational AI, being an English native speaker and being at least 18 years old. In total, the survey was started by 611 participants. After checking the participation requirements, which were clearly stated in the invitation to the survey, 281 seemingly matching participants were forwarded to the actual survey. 20 of them were excluded from analysis due to ambiguous statements about their

conversational AI usage, whereas 256 completed the survey with consent to use data. As pre-registered, we additionally excluded three participants who (1) did not complete the survey using a computer or tablet, (2) did not pass both attention checks, (3) did not consider their own data useable, and/or (4) had an *SDR* > 2.59 in the regression analyses testing the main prediction. Thus, the eligible sample consisted of $N = 253$ participants (as reported in the manuscript).

Pre-Registration

Created:	09/14/2021 02:00 AM (PT)
Made public:	11/28/2022 01:17 AM (PT)
Available at:	https://aspredicted.org/39en3.pdf

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

H1: Competencies on the intellectual level predict a higher perceived warmth.

H2: Competencies on the intellectual level predict (a) a higher willingness to disclose information and (b) lower privacy concerns.

H3: The effect of competencies on the intellectual level on (a) willingness to disclose information and (b) privacy concerns is mediated via perceived warmth.

--

For competencies on the level of meta-cognitive heuristics, different predictions can be derived. In line with the preregistration of our previous study and the results of SEM we would expect:

H4a: Competencies on the level of meta-cognitive heuristics predict a lower perceived warmth.

H5a: Competencies on the level of meta-cognitive heuristics predict (a) a lower willingness to disclose information and (b) higher privacy concerns.

However, based on the idea that warmth requires higher order competencies as well as on bivariate correlations from previous studies, we would predict:

H4b: Competencies on the level of meta-cognitive heuristics predict a higher perceived warmth.

H5b: Competencies on the level of meta-cognitive heuristics predict (a) a higher willingness to disclose information and (b) lower privacy concerns.

In any case we predict:

H6: The effect of competencies on the level of meta-cognitive heuristics on (a) willingness to disclose information and (b) privacy concerns is mediated via perceived warmth.

3) Describe the key dependent variable(s) specifying how they will be measured.

- *Warmth (5 items)*
- *Privacy concerns (5 items)*
- *Willingness to disclose information (5 items)*

4) How many and which conditions will participants be assigned to?

We will implement a 2 (intellectual competencies: low vs high) x 2 (meta-cognitive competencies: low vs high) between-subjects design. Participants will be randomly assigned to one of the four conditions implemented via short descriptions of a technology.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will conduct a MANOVA with the two experimental factors as independent variables and warmth, privacy concerns, and information disclosure as dependent variables. Main effects in the univariate tests will be used to test Hypotheses 1, 2, 4 & 5. Hypotheses 3 & 6 will be tested with a path model including the respective mediation model and the other experimental factor as well as the interaction between both factors as covariates. These tests are conducted using PROCESS model 4 (Hayes, 2018). The indirect effects in these models are used to test these hypotheses via bootstrapped confidence intervals.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Inclusion requirements for participants:

- *No regular use of voice assistants,*
- *at least 18 years old,*
- *consent for participation and usage of data for scientific purposes,*
- *survey completion using computer or tablet,*
- *passing two attention checks.*

Participants will be excluded if they state that they did not answer all questions honestly and attentively. After excluding participants according to these criteria, data will be checked for outliers using studentized deleted residuals (SDR) from a multiple regression analysis of information disclosure on warmth, the two experimental factors,

and their interaction. Participants with an absolute SDR > 2.59 will be regarded as statistical outliers.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.

We seek for a power of 80% and an alpha error of 0.05 for observing a minimum effect of f = 0.15 in an ANOVA with four groups. Hence, we aim to collect N = 351 valid cases. To account for data exclusion based on the above described criteria, we will collect data from N = 360 participants.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will further assess perceived competencies on the intellectual level (3 items) and on the level of meta-cognitive heuristics (3 items).

Deviations from Pre-Registration

Instead of the term 'voice assistant' used in the pre-registration, we refer to 'conversational AI' in the paper. Besides, we pre-registered hypotheses on the association between intellectual competencies and meta-cognitive heuristics with warmth (H1, H4) and potential mediations (H3, H6), which are not reported in the manuscript. Thus, we also pre-registered a MANOVA with the two experimental factors as independent variables and warmth, privacy concerns, and information disclosure as dependent variables – in the manuscript we report the respective MANOVA excluding warmth. The pre-registered MANOVA including warmth as dependent variable indicated no significant multivariate effect of the intellect manipulation, $p = .844$, nor the meta manipulation, $p = .068$, nor the interaction of the two experimental factors, $p = .436$.

We further pre-registered to test several mediation models using PROCESS (model 4; Hayes, 2018). We first conducted mediation analyses with the intellect manipulation as predictor, the meta manipulation and the interaction between the two experimental factors as covariates, warmth as mediator, and (a) willingness to disclose and (b) privacy concerns and as outcomes. We did not observe a total effect of the intellect manipulation on neither willingness to disclose, $p = .506$, nor privacy concerns, $p = .907$. After entering the mediator into the model, the intellect manipulation did not predict perceived warmth (mediator), $p = .637$, which, however, predicted willingness to disclose, $b = 0.55$, $SE = 0.07$, $p < .001$, and privacy concerns, $b = -0.46$. $SE = 0.06$, $p < .001$. The indirect effects were non-significant, for willingness to disclose, 95%-CI [-0.10, 0.06] and privacy concerns, 95%-CI [-0.05, 0.09].

Besides, we conducted mediation analyses with the meta manipulation as predictor, the intellect manipulation and the interaction between the two experimental factors as covariates, warmth as mediator, and (a) willingness to disclose and (b) privacy concerns and as outcomes. We observed a total effect of the meta manipulation on willingness to disclose, $b = 0.23$, $SE = 0.09$, $p = .014$, and privacy concerns, $b = 0.22$, $SE = 0.09$, $p = .013$. After entering the mediator into the model, the meta manipulation did not predict perceived warmth (mediator), $p = .375$, which, however, predicted willingness to disclose, $b = 0.55$, $SE = 0.07$, $p < .001$, and privacy concerns, $b = -0.46$. $SE = 0.06$, $p < .001$. The indirect effects were non-significant, for willingness to disclose, 95%-CI [-0.13, 0.05] and privacy concerns, 95%-CI [-0.38, 0.11].

Additional Analyses

We additionally report bivariate correlations between the manipulation checks (competence levels) and disclosure (see Table A20).

Table A20

Bivariate Correlations Between Competence Levels and Disclosure (Study 1.4: N = 253)

	1	**2**	**3**	**4**
1 Intellectual level (MC)	1.00	0.59	-0.07	0.13
2 Meta-cognitive heuristics (MC)	0.59	1.00	-0.13	0.23
3 Willingness to disclose	-0.07	-0.13	1.00	-0.76
4 Privacy concerns	0.13	0.23	-0.76	1.00

Materials of Study 1.4

Procedure (core measures in bold):

1. Informed consent
2. Introductory text and scenario description (see below)
3. Manipulation of intellectual competencies and meta-cognitive heuristics (see below)
4. **Privacy concerns and willingness to disclose** (10 items, same as in Study 1.2), attention check (1 item)
5. Perceived warmth (5 items)
6. **Manipulation check** (6 items, see below), attention check (1 item)
7. Demographics, debriefing, final consent, option for comment

Introductory text and scenario description:

> *Imagine the following situation: one of your friends recently bought a new smart speaker with an integrated voice assistant for their home. This voice assistant*

communicates with the user via voice by processing voice input and producing voice output.

The voice assistant is able to react to plenty of different user requests, meaning that it is very well prepared to deal with specific situations and tasks by following predefined rules. For example, it can be used to assess information from the internet, to set alarms, to stream media, or to control other smart devices. For this purpose, the voice assistant uses information that the user explicitly shares with it – for example, when the user requests the completion of specific tasks or explicitly submits information about personal preferences.

(page break)

Your friend now tells you about their personal experience with the voice assistant:

"I use the voice assistant a lot when I am at home. I noticed that the voice assistant behaves differently depending on the specific situation: it considers which task I want it to solve, whether it is me or somebody else it is interacting with, when and where the interaction takes place, and what happens during the interaction. It plays different songs when I ask it to play music in the morning compared to when my partner asks or when I ask it to play music in the evening. Then again, when I ask 'What can I do on the weekend?', it gives me information about local events in my neighborhood. Thus, I have the impression that the voice assistant adapts its behavior depending on several factors."

<u>Manipulation of intellectual competencies and meta-cognitive heuristics:</u>

Low intellectual competencies, low meta-cognitive heuristics:

"However, the voice assistant is incapable of dealing with incomplete information and it cannot anticipate potential problems or find innovative solutions. For example, it told me about an interesting open-air event in the city. I decided to go there, but then it rained the whole evening. The voice assistant neither warned me about the rainy weather in advance nor did it propose any alternative indoor events."

(page break)

"In addition, I realized that the voice assistant does not learn from mistakes and cannot use information from prior events to solve related tasks. The voice assistant is, thus, incapable to develop abstract strategies based on previous incidents.

For example, although I have now used the voice assistant for some time, it has neither learned about my personal preferences for events nor the money I am willing to spend.

It still proposes a wide range of events. And some of them are way more costly than I am willing to pay or do not fit my interests.

The voice assistant is also far from improving its other proposals to fit my financial requirements. Just yesterday, for the first time, I asked the voice assistant for advice where to spend my next vacation. Its suggestions were far off my budget – it obviously did not use the information about my budget for other things like events to estimate my travel budget. It looks like the voice assistant cannot improve its suggestion strategy without my explicit input."

High intellectual competencies, low meta-cognitive heuristics:

"Moreover, the voice assistant is capable of dealing with incomplete information and it can also anticipate potential problems and find innovative solutions.

For example, it told me about an interesting open-air event in the city. However, it warned me that the weather forecast announced rainy weather. It advised me to wear proper clothes if I wanted to go to the event and additionally proposed alternative indoor events."

(page break)

"However, I realized that the voice assistant does not learn from mistakes and cannot use information from prior events to solve related tasks. The voice assistant is, thus, incapable to develop abstract strategies based on previous incidents.

For example, although I have now used the voice assistant for some time, it has neither learned about my personal preferences for events nor the money I am willing to spend. It still proposes a wide range of events. And some of them are way more costly than I am willing to pay or do not fit my interests.

The voice assistant is also far from improving its other proposals to fit my financial requirements. Just yesterday, for the first time, I asked the voice assistant for advice where to spend my next vacation. Its suggestions were far off my budget – it obviously did not use the information about my budget for other things like events to estimate my travel budget. It looks like the voice assistant cannot improve its suggestion strategy without my explicit input."

Low intellectual competencies, high meta-cognitive heuristics:

"However, the voice assistant is incapable of dealing with incomplete information and it cannot anticipate potential problems or find innovative solutions. For example, it told me about an interesting open-air event in the city. I decided to go there, but then

it rained the whole evening. The voice assistant neither warned me about the rainy weather in advance nor did it propose any alternative indoor events."
(page break)
"However, I realized that the voice assistant learns from mistakes and can use information from prior events to solve related tasks. The voice assistant is, thus, capable to develop abstract strategies based on previous incidents.
For example, after I had used the voice assistant for some time, it obviously learned about my personal preferences for events and about the amount of money I am willing to spend. It now proposes me fewer events, but the proposed one's fit much better to my interests and my budget.
It also seems that the voice assistant's other proposals do now better fit my financial requirements. Just yesterday, for the first time, I asked the voice assistant for advice where to spend my next vacation. It obviously chose destinations that I can afford - it must have used the information about my budget for other things like events to estimate my travel budget. It looks like the voice assistant improved its strategy for suggestions without the need for my explicit input."

High intellectual competencies, high meta-cognitive heuristics:

"Moreover, the voice assistant is capable of dealing with incomplete information and it can also anticipate potential problems and find innovative solutions. For example, it told me about an interesting open-air event in the city. However, it warned me that the weather forecast announced rainy weather. It advised me to wear proper clothes if I wanted to go to the event and additionally proposed alternative indoor events."
(page break)
"In addition, I realized that the voice assistant learns from mistakes and can use information from prior events to solve related tasks. The voice assistant is, thus, capable to develop abstract strategies based on previous incidents.
For example, after I had used the voice assistant for some time, it obviously learned about my personal preferences for events and about the amount of money I am willing to spend. It now proposes me fewer events, but the proposed one's fit much better to my interests and my budget.
It also seems that the voice assistant's other proposals do now better fit my financial requirements. Just yesterday, for the first time, I asked the voice assistant for advice where to spend my next vacation. It obviously chose destinations that I can afford – it must have used the information about my budget for other things like events to estimate

my travel budget. It looks like the voice assistant improved its strategy for suggestions without the need for my explicit input."

Manipulation check:

Intellectual competencies:

- *The voice assistant solves unknown tasks.*
- *The voice assistant can plan actions.*
- *The voice assistant can extract key information from complex input.*

Meta-cognitive heuristics:

- *The voice assistant adapts its behavior based on prior events.*
- *The voice assistant transfers knowledge to other domains.*
- *The voice assistant develops general strategies to accomplish various tasks.*

Additional Information for Study 1.5

Exclusion of Participants

As pre-registered, we applied the same Prolific pre-screeners as in Study 1.3. The survey was started by 572 participants and completed with consent to use data by 565 participants of which we excluded 46 based on the same pre-registered criteria as in Study 1.3 (20 did not complete the survey using a computer or tablet, 21 did not pass both attention checks, five did not consider their own data useable) and, additionally, 48 participants indicating to be users (also pre-registered for this study). Thus, the eligible sample consisted of $N = 471$ participants (as reported in the manuscript).

Pre-Registration

Created:	02/02/2022 05:18 AM (PT)
Made public:	11/28/2022 01:19 AM (PT)
Available at:	https://aspredicted.org/947b9.pdf

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

H1: Higher competencies on the intellectual level lead to (a) higher willingness to disclose information and (b) lower privacy concerns.

3) Describe the key dependent variable(s) specifying how they will be measured.

- *Privacy concerns (5 items)*
- *Willingness to disclose information (5 items)*

4) How many and which conditions will participants be assigned to?

We will implement a one factorial between-subjects design (competencies on the intellectual level: low / high). Participants will be randomly assigned to one of the two resulting conditions implemented via short descriptions of a CAI.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will run a MANOVA with the experimental factor as independent variable and privacy concerns as well as willingness to disclose information as dependent variables. Support for Hypotheses 1 a & b will be provided by the separate tests for the respective dependent variable.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Inclusion requirements for participants:

- *at least 18 years old,*
- *consent for participation and usage of data for scientific purposes,*
- *survey completion using computer or tablet,*
- *passing two attention checks,*
- *being a non-user of CAIs (i.e., using a CAI maximum once a month).*

Participants will be excluded if they state that they did not answer all questions honestly and attentively.

After excluding participants according to these criteria, data will be checked for outliers using studentized deleted residuals (SDR) from a multiple regression analysis of willingness to disclose information on the experimental factor. Participants with an absolute SDR > 2.59 will be regarded as statistical outliers.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We seek for a power of 80% and an alpha error of 0.05 for observing a minimum effect of f2 = 0.02 in a MANOVA with two groups, one independent and two dependent variables. Hence, we aim to collect N = 486 valid cases. To account for data exclusion based on the above-described criteria, we will collect data from N = 535 participants (i.e., oversampling of 10%).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will assess perceived competencies on the intellectual level (3 items) and on the level of meta-cognitive heuristics (3 items) as a manipulation check.

Deviations from Pre-Registration & Additional Analyses

Instead of the abbreviation 'CAI' used in the pre-registration, we refer to 'conversational AI' in the paper. We additionally report bivariate correlations between competence levels and disclosure (see Table A21).

Table A21

Bivariate Correlations Between Competence Levels and Disclosure (Study 1.5: N = 471)

	1	**2**	**3**	**4**
1 Intellectual level (MC)	1.00	0.52	0.13	-0.03
2 Meta-cognitive heuristics (MC)	0.52	1.00	0.01	0.10
3 Willingness to disclose	0.13	0.01	1.00	-0.70
4 Privacy concerns	-0.03	0.10	-0.70	1.00

Materials of Study 1.5

Procedure (core measures in bold):

1. Informed consent
2. Introductory text and scenario description (see below)
3. Manipulation of intellectual competencies (see below)
4. **Privacy concerns and willingness to disclose** (10 items, same as in Study 1.2), attention check (1 item)
5. **Manipulation check** (6 items, see below), attention check (1 item)
6. Usage (2 items)
7. Demographics, debriefing, final consent, option for comment

Introductory text and scenario description:

Imagine the following situation: one of your friends recently bought a new smart speaker with an integrated voice assistant for their home. This voice assistant communicates with the user via voice by processing voice input and producing voice output.

The voice assistant is able to react to plenty of different user requests, meaning that it is very well prepared to deal with specific situations and tasks by following predefined

rules. For example, it can be used to assess information from the internet, to set alarms, to stream media, or to control other smart devices. For this purpose, the voice assistant uses information that the user explicitly shares with it – for example, when the user requests the completion of specific tasks or explicitly submits information about personal preferences.

(page break)

Your friend now tells you about their personal experience with the voice assistant: 'I use the voice assistant a lot when I am at home. I noticed that the voice assistant behaves differently depending on the specific situation: it considers which task I want it to solve, whether it is me or somebody else it is interacting with, when and where the interaction takes place, and what happens during the interaction. It plays different songs when I ask it to play music in the morning compared to when my partner asks or when I ask it to play music in the evening. Then again, when I ask 'What can I do on the weekend?', it gives me information about local events in my neighborhood. Thus, I have the impression that the voice assistant adapts its behavior depending on several factors.'

Manipulation of intellectual competencies:

Low intellectual competencies:

'However, the voice assistant is not able to gather key information out of complex input. It also does not plan actions in advance, nor does it find innovative solutions.

For example, some days ago, I asked the voice assistant to read out an email in which I was invited to a work meeting in another city. As I decided to go there, I asked the voice assistant how to get there. The voice assistant did not show me different options, but I had to specify which means of transportation I preferred for this journey. When I said that I want to go by car, it suggested to start in the middle of the night to be there in time instead of leaving the day before and spend the night in a hotel. Further, I had to manually add the appointment and my estimated travel time to my calendar afterward.'

High intellectual competencies:

c

Manipulation check:

Intellectual competencies:

- *The voice assistant anticipates potential problems.*

- *The voice assistant can deal with incomplete information,*
- *The voice assistant solves unknown tasks.*

Meta-cognitive heuristics:

- *The voice assistant adapts its behavior based on prior events.*
- *The voice assistant transfers knowledge to other domains.*
- *Based on previous incidents, the voice assistant derives universal strategies.*

Appendix B: Additional Information for Chapter 3

As pointed out in Chapter 3, the data collection of Studies 2.1 and 2.2 were part of a larger project. This larger project is described in Chapter 2 of the current dissertation. Hence, data for Study 2.1 was collected together with Study 1.1 and data for Study 2.2 was collected together with Study 1.2. Accordingly, information on exclusion of participants, data set, (deviations from) pre-registration, and procedure can be found in Appendix A.

Appendix C: Additional Information for Chapter 4

Data set and the analysis syntax are available for peer review on researchbox.org: https://researchbox.org/1574&PEER_REVIEW_passcode=VWROJJ. They will be made publicly available for scientific use after the manuscript has been accepted for publication.

Pre-Registration

Created:	03/21/2023 01:05 AM (PT)
Made public:	Not yet, will be made publicly available after publication
Available at:	https://aspredicted.org/Y9F_TKV (anonymized, for peer review)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

H1: Participants share less information if the system provider is less (vs more) trustworthy.

H2: The negative effect of low provider trustworthiness will be weaker when participants perceive the system output as more qualitative.

3) Describe the key dependent variable(s) specifying how they will be measured.

- *Perceived output quality (4 items)*
- *Information disclosure (Number of information that participants are willing to share from a list of 20 items)*

4) How many and which conditions will participants be assigned to?

We will implement a one-factorial between-subjects design. Participants will be randomly assigned to one of the following conditions: (1) low provider trustworthiness, (2) high provider trustworthiness.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will conduct a multiple regression analysis regressing information disclosure on perceived output quality, provider trustworthiness, and the interaction of these two factors. Evidence for H1 will be provided by a significant main effect of provider trustworthiness, evidence for H2 will be provided by a significant interaction.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will include participants for analysis if they pass two attention checks and are fluent in German (the language the study is conducted in).

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We seek for a power of 80% and an alpha error of 0.05 for observing a minimum effect of f2 = 0.03 in a multiple regression with 3 predictors. Hence, we aim to collect N = 368 valid cases. To account for data exclusion based on the above-described criteria, we aim to collect N = 405 cases (i.e., oversampling of 10%).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will further assess:

- *Privacy concerns (5 items)*
- *Intention to use (1 item)*
- *Perceived usefulness (1 item)*
- *Perceived ease of use (1 item)*
- *Manipulation check: trust in provider (6 items)*
- *Individual relevance of breakfast, internet recipes and cooking (1 item each)*

Additional Analyses

Descriptive Results for Willingness to Disclose

Table C1

Percentage of participants willing to disclose specific types of information in Study 3 (N = 329)

Type of information	Percentage willing to disclose
Information, which groceries I (dis)like	94.5%
Age	72.6%
Gender	72.3%
Fitness activities (duration and intensity per week)	69.0%
Weight	66.9%
Body height	65.0%
Favorite (grocery) shops	62.9%
Weekly budget for groceries	61.7%
Frequency of grocery shopping	59.0%
Preferred means of transport (e.g., for getting to work)	47.1%
Information about diseases	45.3%
E-mail address	41.0%
Relationship and family status	31.9%
Location	15.2%
IP address	10.3%
Income	10.0%
Postal address	7.0%
Connection to other data sources (e.g., Amazon, Payback)	6.7%
Phone number	4.3%
Browser history	3.6%

Note. Participants were asked: 'Which of the following data would you share with the recipe guide to get better-fitting recommendations?'

Descriptive Results and MANOVA for Single Items of Perceived Provider Trustworthiness (Manipulation Check)

As stated in the manuscript, we assessed perceived provider trustworthiness (i.e., the manipulation check) with six items covering benevolence, integrity, and competence. As an exploratory factor analysis indicated a one-factor solution we summarized these six items into a scale and used this scale for the manipulation check reported in the manuscript.

We further exploratorily conducted a MANOVA to assess differences between the experimental conditions (low vs. high provider trustworthiness, as IV) regarding the single items of perceived trustworthiness (as DVs). This analysis as well as the descriptive results

(both reported in Table C2) indicate that the perceptions of the two providers differed in benevolence and integrity but not regarding their competence to provide a recipe guide.

Table C2

Descriptive Results and MANOVA Results for Single Items of Perceived Provider Trustworthiness (Manipulation Check)

Item	Low trustworthiness condition (N = 174)		High trustworthiness condition (N = 155)		MANOVA	
	M	*SD*	*M*	*SD*	$F(1, 327)$	*p*
Benevolence item 1	4.07	1.48	5.19	1.17	56.26	<.001
Benevolence item 2	4.21	1.48	5.10	1.17	35.35	<.001
Integrity item 1	4.34	1.30	5.15	1.16	34.83	<.001
Integrity item 2	4.17	1.56	5.10	1.06	39.02	<.001
Competence item 1	5.12	1.25	5.22	1.18	0.54	0.465
Competence item 2	5.06	1.41	5.18	1.27	0.62	0.430

Individual Relevance of Cooking and Breakfast as Additional Predictor

We additionally tested whether the results for willingness to use are moderated by the induvial relevance of cooking and breakfast (assessed with three items, e.g., ‘ I frequently use recipes from the internet’, ‘A diverse breakfast is important for me’, translated from German, on a seven-point scale ranging from 1 = *strongly disagree* to 7 = *strongly agree,* $\alpha = 0.72$, $M = 4.82$, $SD = 1.24$). As displayed in Table C3, the individual relevance of cooking and breakfast was indeed a strong predictor of intention to use but did neither interact with the other factors nor change the interpretation of the respective results reported in the manuscript.

Table C3

Results of Regression Analysis for Intention to Use Including Individual Relevance of Cooking and Breakfast as Additional Predictor

Predictor	*B*	*SE*	β	$t(321)$	*p*
Output quality	**0.86**	**0.06**	**0.58**	**13.64**	**< .001**
Provider trustworthiness	0.03	0.06	0.02	0.50	.618
Output quality x Provider trustworthiness	0.06	0.06	0.04	0.97	.333
Personal relevance	**0.41**	**0.06**	**0.28**	**6.65**	**< .001**
Output quality x Personal relevance	-0.01	0.05	-0.01	-0.15	.883
Provider trustworthiness x Personal relevance	-0.03	0.06	-0.02	-0.44	.660
Output quality x Provider trustworthiness x Personal relevance	-0.02	0.05	-0.02	-0.39	.693

Materials of Study 3

Procedure (core measures in bold):

1. Informed consent & prescreening
2. Information about provider (depending on experimental condition)

3. Interaction with experimentally manipulated recipe guide
4. **Perceived output quality** (4 items, see below)
5. **Willingness to disclose** (20 items, see Table C1)
6. Privacy concerns (5 items), attention check (1 item)
7. Usage intention, usefulness, ease of use (1 item each)
8. **Manipulation check** (trust in provider, 6 items, see below), attention check (1 item)
9. Individual relevance of cooking and breakfast (3 items)
10. Demographics, debriefing, final consent, option for comment

Measures (Original German items and English translation):

Perceived output quality:

- *Ich bin mit den vorgeschlagenen Rezepten zufrieden. [I am satisfied with the recommended recipes.]*
- *Die Rezeptauswahl war für mich passend. [The recipe recommendations were appropriate for me.]*
- *Ich fand die Rezeptauswahl hilfreich. [I found the recipe recommendations helpful.]*
- *Die Rezeptvorschläge klingen für mich lecker. [The recipe recommendations sound delicious to me.]*

Manipulation check:

- *Der Anbieter ist interessiert am Wohl der Nutzenden des Rezept-Assistenten. [The provider is interested in the well-being of the recipe guide.]*
- *Ich bin überzeugt, dass der Anbieter im Interesse der Nutzenden handelt.[I am convinced that the provider acts in the best interest of the users.]*
- *Der Anbieter ist ehrlich im Umgang mit den Nutzenden. [The provider is honest in its dealing with the users.]*
- *Der Anbieter ist aufrichtig. [The provider is sincere.]*
- *Der Anbieter ist kompetent für die Bereitstellung eines Rezept-Assistenten. [The provider is competent for providing a recipe guide.]*
- *Der Anbieter ist fähig, gute Rezept-Empfehlungen zur Verfügung zu stellen. [The provider is capable of providing good recipe recommendations.]*

Appendix D: Additional Information for Chapter 5

Data set and analysis syntax are available for peer review on researchbox.org: https://researchbox.org/1666&PEER_REVIEW_passcode=ZEGPZW.

Additional Information for Study 4.1

Exclusion of Participants

In total, the survey was started 265 times. We excluded submissions from participants who started the survey more than once (9 cases). As pre-registered, we further excluded participants without management experience (69 cases), who were not fluent in German (7 cases), or who did not pass the attention checks (10 cases). We further excluded two cases with implausible age values (i.e., 3 and 5 years). Thus, the eligible sample consisted of $N = 168$ participants (as reported in the manuscript).

Pre-Registration

Created:	09/19/2022 11:55 PM (PT)
Made public:	Not yet, will be made publicly available after publication
Available at:	https://aspredicted.org/VQS_2NX (anonymized, for peer review)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

H1: Usage of AI in companies is more acceptable in the area of finances (vs the area of human resources).

RQ1: Is the effect of application area on acceptance (described in H1) moderated by task stage?

3) Describe the key dependent variable(s) specifying how they will be measured.

Acceptance:

- *Benefit (1 item)*
- *Potential to handle future challenges (1 item)*
- *Potential to heighten success (1 item)*
- *Risk of negative consequences (1 item)*
- *Willingness to invest (1 item)*

4) How many and which conditions will participants be assigned to?

We will implement a two-factorial mixed-measures design. Participants will be randomly assigned to one potential application area of AI in the business context

(between-factor: finances vs human resources). Each participant will then be exposed to four consecutive task stages of information processing reflecting increasing autonomy of the AI (within-factor: 1. Monitoring, 2. Generating options, 3. Selection, 4. Implementation). The order of the tasks will not be randomized.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

In case that the five measurements of acceptance (see point 3) form a scale with α > .7, we will average them into one index of acceptance and submit it to the analysis reported below. Otherwise, we will conduct the following analysis separately for each of the single-item measures of acceptance.

We will calculate a multi-level model regressing acceptance on application area and task stage (assuming a random intercept for participants). Application area (finances vs human resources) will be included as level-2 predictor (fixed effect). Task stage (monitoring, generating options, selection, implementation) will be included as level-1 predictor (fixed effect) by using contrast coding ([c1] 0.5, -0.5, 0, 0; [c2] 0, 0, 0.5. -0.5, [c3] 0.25, 0.25, -0.25, -0.25). We will further include the interaction between application area and task stage. Evidence for H1 will be provided by a main effect of application area, whereas evidence for RQ1 will be provided by interaction effects between application area and task stage.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Inclusion requirements for participants:

- *Experience being in a management position*
- *Fluent in German (language the study is conducted in)*
- *Passing two attention checks and stating two have answered all questions attentively*

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

As participants matching our inclusion criteria are difficult to recruit, we aim to sample N = 150 valid cases. To account for data exclusion based on the above-described criteria, we will collect data from N = 175 participants (i.e., oversampling of appr. 15%). If no additional submissions come in within 72 hours while we have not reached our desired sample size, we will stop the first wave of data collection and resend the

study invitation to suitable potential participants who did not participate in the study by then. After another 72 hours without new submissions, we will stop data collection.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will further assess perceived competencies of existing AI tools (24 items) and innovative behavior (8 Items) as well as for each task stage usefulness (4 items), perceived autonomy, controllability, and task complexity (1 item each).

Deviations from Pre-Registration

We slightly adjusted the wording of the preregistration in the manuscript to be consistent across studies (see Table D1 for all changes in the wording of Study 4.1).

Table D1

Changes in Wording of Study 4.1

Change	Wording in preregistration	Wording in paper
1	application area	business area
2	task stage	functionality
3	generating options	generation
H1	Usage of AI in companies is more acceptable in the area of finances (vs the area of human resources).	Usage of AI is less accepted in HR than in other business areas (i.e., finances, marketing).
RQ	Is the effect of application area on acceptance (described in H1) moderated by task stage?	Does functionality influence AI acceptance and does this effect depend on the business area?

We further pre-registered to regress acceptance on business area and functionality (assuming a random intercept for participants) in case the five measurements (1) benefit, (2) potential to handle future challenges, (3) potential to heighten success, (4) risk of negative consequences, and (5) willingness to invest form a scale with $\alpha > .7$. The reliability analysis showed a Cronbach's alpha = 0.82 for this scale. The following multi-level model was pre-registered and is reported in Table D2: business area (finances vs. human resources) was included as a level-2 predictor (fixed effect), functionality (monitoring, generation, selection, implementation) as a level-1 predictor (fixed effect) by using contrast coding ([c1] 0.5, -0.5, 0, 0; [c2] 0, 0, 0.5. -0.5, [c3] 0.25, 0.25, -0.25, -0.25), besides the interaction between business area and functionality was included. We refrained from reporting this analysis in the paper to use consistent analyses across studies.

Table D2

Multi-Level Model Regressing Acceptance on Business Area and Functionality (Study 4.1: N = 168)

Effect	*Estimate*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	4.66	0.07	166	69.59	<.001
Business area[a]	0.18	0.07	166	2.74	.007
Functionality.C1[b]	-0.05	0.07	498	-0.73	.468
Functionality.C2[c]	0.43	0.07	498	6.15	<.001
Functionality.C3[d]	0.57	0.10	498	5.70	<.001
Business area x Functionality.C1	0.19	0.07	498	2.76	.006
Business area x Functionality.C2	-0.01	0.07	498	-0.10	.919
Business area x Functionality.C3	0.09	0.10	498	0.87	.383
Random effects	Variance	SD			
Participant	0.65	0.81			
Residual	0.42	0.65			

[a] human resources = -1, finances = 1.

[b] monitoring = 0.5, generation = -0.5, selection/implementation = 0.

[c] monitoring/generation = 0, selection = 0.5, implementation – 0.5.

[d] monitoring/generation = 0.25, selection/implementation = -0.25.

Materials of Study 4.1

Procedure of Study 4.1

1. Informed consent and prescreening
2. Perceived competencies of existing AI tools (24 items), attention check (1 item)
3. Description of AI functionalities in HR/Finances (depending on experimental condition, see below)
4. Acceptance (5 items per AI functionality, including **perceived risk of negative consequences and willingness to invest**, see manuscript), perceived usefulness (4 items per AI functionality), perceived autonomy of the AI, perceived human control over the AI, perceived task complexity (1 item each per AI functionality)
5. Innovative behavior (8 items), attention check (1 item)
6. Demographics, final consent, option for comment

Description of AI functionalities in HR [Finances]

Below the original German materials as well as English translations are provided. The underlined text parts are condition-specific and were displayed either in the human resources condition or in the finances condition (the latter being provided in square brackets).

Introduction (German):

> *Im weiteren Fragebogen geht es um spezifische Einsatzmöglichkeiten von KI in Unternehmen. Dazu werden Sie gleich ein Anwendungsbeispiel lesen, in dem KI im Personalwesen zur Ansprache von potenziellen Bewerbern und Bewerberinnen [im Controlling zur Optimierung der Verteilung von Marketingausgaben auf unterschiedliche Kanäle] genutzt wird. In diesem Beispiel werden vier aufeinander folgende Teilschritte dieses Prozesses von der KI übernommen. Die Teilschritte werden Ihnen nacheinander präsentiert, danach werden Sie jeweils einige Fragen zur Anwendung von KI für den jeweiligen Teilschritt beantworten.*
>
> *In diesem Fallbeispiel geht es um ein Unternehmen, das mehrere, ähnliche offene Stellen besetzen möchte. Um die Anzahl der geeigneten Bewerbungen zu erhöhen, sollen neue Wege gegangen werden und KI eingesetzt werden, um die Bewerberansprache zu optimieren [seine monatlichen Umsatzüberschüsse den unterschiedlichen Marketingkanäle zuordnen möchte. Um möglichst effiziente Investitionsstrategien zu verfolgen, sollen neue Wege gegangen und KI zur Strategieauswahl eingesetzt werden].*

Introduction (Translated):

> *The following questionnaire deals with specific scenarios of AI usage in companies. For this purpose, you will read an example of AI usage in human resources to contact potential applicants [in controlling to optimize the distribution of marketing expenses across different channels]. In this example, four successive sub-steps of this process are taken over by AI. The sub-steps will be presented to you one after the other. After each sub-step, you will answer some questions about the application of AI for the respective sub-step.*
>
> *This case study is about a company that wants to fill several, similar open positions. In order to increase the number of suitable applications, the company wants to use new approaches and use AI to optimize applicant targeting [allocate its monthly revenue surpluses to the different marketing channels. To pursue the most efficient investment strategies, the company wants to use new approaches and use AI for strategy selection].*

Monitoring (German):

Die KI wird zunächst eingesetzt, um Daten aus vorherigen Bewerbungsprozessen auszuwerten, in denen vergleichbare Stellen besetzt wurden. Hierbei identifiziert die KI relevante Eigenschaften und Fähigkeiten bisheriger erfolgreicher Stelleninhaber und Stelleninhaberinnen [historische Marketing-Performancedaten auszuwerten, die Rückschlüsse auf die Effizienz einzelner Marketing-Kanäle erlauben. Hierbei identifiziert die KI relevante Kanäle, die zur Verfolgung von effizienten Investitionsstrategien genutzt werden können].

Monitoring (Translated):

The AI is initially used to evaluate data from previous application processes in which comparable positions were filled. Here, the AI identifies relevant characteristics and skills of previous successful job holders [historical marketing performance data that allows conclusions about the efficiency of individual marketing channels. Here, the AI identifies relevant channels that can be used to pursue efficient investment strategies].

Generation (German):

Aufbauend auf den Ergebnissen der Datenanalyse schlägt die KI nun verschiedene Kanäle vor, deren Nutzung für die Rekrutierung von Personen mit den gesuchten Eigenschaften und Fähigkeiten vielversprechend sind [Investitionsstrategien vor, welche den zu investierenden Betrag effizient auf die unterschiedlichen Kanäle aufteilen würden].

Generation (Translated):

Building on the results of the data analysis, the AI now suggests different channels that are promising for recruiting individuals with the required characteristics and skills [investment strategies that would efficiently allocate the available amount among the different channels].

Selection (German):

Im nächsten Schritt wählt die KI aus den vorgeschlagenen Kanälen den geeignetsten Kanal (oder die geeignetste Kombination aus Kanälen) für die Besetzung der offenen Stellen aus [Investitionsstrategien die geeignetste Strategie (oder Kombination aus Strategien) für die Nutzung des vorhandenen Investitionsbudgets aus].

Selection (Translated):

> *In the next step, the AI selects the most suitable channel (or the most suitable combination of channels) for filling the vacancies [investment strategy (or combination of strategies) for using the available investment budget].*

Implementation (German):

> *Im letzten Schritt übernimmt die KI auch die Generierung und Veröffentlichungen von Inhalten auf den ausgewählten Kanälen: die KI erstellt und veröffentlich also beispielsweise eigenständig Stellenausschreibungen oder Social Media Posts [Investitionsanweisung an die unterschiedlichen Marketing-Teams: die KI informiert die jeweiligen Kanalverantwortlichen über das Budget, welches dem Kanal im kommenden Monat zur Verfügung steht].*

Implementation (Translation):

> *In the last step, the AI also takes over the generation and publication of content on the selected channels: for example, the AI independently creates and publishes job advertisements or social media posts [the investment instructions to the various marketing teams: the AI informs the respective channel managers about the budget available to the channel in the coming month].*

Additional Information for Study 4.2a

Exclusion of Participants

In total, the survey was started 541 times. As pre-registered, we excluded participants without management experience (35 cases), working less than 19 hours a week (30 cases), failing at least one attention check (22 cases), or not changing the default value for any of the acceptance measures (3 cases). We further excluded three participants who were not fluent in English (the language the study was conducted in). Thus, the eligible sample consisted of $N = 451$ participants (as reported in the manuscript).

Pre-Registration

Created:	12/05/2022 02:37 AM (PT)
Made public:	Not yet, will be made publicly available after publication
Available at:	https://aspredicted.org/VF6_L74 (anonymized, for peer review)

1) Have any data been collected for this study already?

> *No, no data have been collected for this study yet.*

2) What's the main question being asked or hypothesis being tested in this study?

H1: Usage of AI in companies is more acceptable in the area of finances (vs the area of human resources).

H2: Usage of AI in companies is more acceptable with lower functionality of the AI.

3) Describe the key dependent variable(s) specifying how they will be measured.

Acceptance:

- *Benefit (1 item)*
- *Potential to increase success (1 item)*
- *Risk of negative consequences (1 item)*
- *Willingness to invest (1 item)*

4) How many and which conditions will participants be assigned to?

We will implement a two-factorial mixed-measures design. Participants will be randomly assigned to one potential application area of AI in the business context (between-factor: finances vs human resources). Each participant will then be exposed to four consecutive subtasks reflecting increasing functionality of the AI (within-factor: 1. Monitoring, 2. Generating, 3. Selecting, 4. Implementing). The order of the subtasks will not be randomized.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

In case that the four measurements of acceptance (see point 3) form a scale with $\alpha > .7$, we will average them into one index of acceptance and submit it to the analysis reported below. Otherwise, we will conduct the following analysis separately for each of the single-item measures of acceptance.

We will calculate a mixed ANOVA with application area (finances vs human resources) as between-factor, functionality as within-factor, and acceptance as dependent variable. In case of a main effect of functionality (monitoring, generating, selecting, implementing), we will use contrast-coding as follows in the follow-up analysis: [c1] -1/4, -1/4, -1/4, 3/4, [c2] -2/3, 1/3, 1/3, 0; [c3] 0, -1/2, 1/2, 0. Evidence for H1 will be provided by a main effect of application area, whereas evidence for H2 will be provided by a main effect of functionality.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Inclusion requirements for participants:

- *Experience being in a management position*

- *Working at least 19 hours a week*
- *Passing two attention checks and stating two have answered all questions attentively.*

We will further exclude participants who do not change the default value (i.e., 0) for any of the acceptance measures (i.e., participants with a 0-value on all four measures of acceptance for all four functionalities will be excluded).

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We seek for a power of 90% and an alpha error of 0.05 for observing a minimum effect of f = 0.15 (based on an earlier study) for the between-subjects factor in the mixed ANOVA described under (5). Hence, we aim to sample N = 400 valid cases. To account for data exclusion based on the above-described criteria, we will collect data from N = 500 participants (i.e., oversampling of appr. 20%).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will further assess the perceived potential to save working hours (1 item) in the respective application area (i.e., finances or human resources).

Deviations from Pre-Registration

We slightly adjusted the wording of the preregistration in the paper to be consistent across studies (see Table D3 for all changes in the wording of Study 4.2a).

Table D3

Changes in Wording of Study 4.2a

Change	Wording in preregistration	Wording in paper
1	application area	business area
H1	Usage of AI in companies is more acceptable in the area of finances (vs the area of human resources).	Usage of AI is less accepted in HR than in other business areas (i.e., finances, marketing).
RQ	Usage of AI in companies is more acceptable with lower functionality of the AI.	Does functionality influence AI acceptance and does this effect depend on the business area?

We pre-registered to average the four measures of acceptance (perceived benefit, perceived potential, perceived risk of negative consequences, willingness to invest) into one index in case they formed a scale with $\alpha > .70$. Although the acceptance scale had an $\alpha = .74$, we report separate analyses for perceived risk and willingness to invest in the manuscript to be

consistent across studies. The analysis results with the preregistered acceptance scale are reported in Table D4.

Table D4

Results of a Mixed ANOVA for Acceptance in Study 4.2a (N = 451)

	Risk of negative consequences			
Predictor	*F*	*df*	*p*	η_p^2
Functionality	17.29	2.86, 1284,47	< .001	0.04
Business area[a]	9.10	1, 449	.003	0.02
Functionality x Business area	19.42	2.86, 1284.47	< .001	0.04
Follow-up analysis: HR				
C1[b]	9.91	1, 225	.002	0.04
C2[c]	25.94	1, 225	< .001	0.10
C3[d]	0.13	1, 225	.717	0.00
Follow-up analysis: Finances				
C1[b]	5.52	1, 224	.020	0.02
C2[c]	8.89	1, 224	.003	0.04
C3[d]	86.79	1, 224	< .001	0.28

Note. In case that sphericity was violated, we report Huynh-Feldt corrected values.

[a] HR = 1, finances = 2.

[b] Monitoring/generation/selection = -1/4, implementation = 3/4.

[c] Monitoring = -2/3, generation/selection = 1/3, implementation = 0.

[d] Monitoring = 0, generation = -1/2, selection = ½, implementation = 0.

Materials of Study 4.2a

Procedure of Study 4.2a

1. Informed consent and prescreening
2. Current use of AI in company (1 item)
3. Potential to save working hours in HR/Finances (depending on experimental condition)
4. Description of AI functionalities in HR/Finances (depending on experimental condition, see below)
5. Acceptance (4 items per AI functionality, including **perceived risk of negative consequences and willingness to invest**, see manuscript)
6. Attention check (1 item)
7. Demographics, final consent, option for comment

Description of AI functionalities in HR [Finances]

Below the original materials are provided. The underlined text parts are condition-specific and were displayed either in the human resources condition or in the finances condition (the latter being provided in square brackets).

Introduction:

> *In the following, you will read a case study of AI being used in the human resources [finances] sector of a company. The AI will take over four consecutive subtasks which will be presented to you in a sequential order. After reading each task description, you will answer some questions about AI for the respective subtask.*
>
> *This case study is about a company that wants to fill several similar vacancies. In order to increase the number of relevant applications, they want to take new approaches and use AI to optimize the recruitment of potential applicants [detect violation of policies regarding the cash flow. In order to increase policy compliance, they want to take new approaches and use AI in the process of preventing policy violations].*

Monitoring:

> *The AI evaluates anonymized application and performance data of current employees in similar positions. In doing so, the AI highlights potentially relevant characteristics of successful employees [past cash flow data of the company. In doing so, the AI highlights potentially relevant indicators of suspicious transactions].*

Generation:

> *Using the results of the data analyses, the AI scans work-related social media profiles (e.g., LinkedIn) and rates the suitability of the respective persons for the vacant position [the current cash flow and rates the suspiciousness of each transaction].*

Selection:

> *Based on the ratings of work-related social media profiles, the AI selects a certain number of persons to be considered in the further recruitment process [about the suspiciousness of transactions, the AI flags highly suspicious transactions].*

Implementation:

> *The AI invites the selected persons to apply for the vacancies. In detail, the AI autonomously messages them that they have been identified as suitable candidates. [decides about the (non-)execution of transactions. In detail, the AI autonomously puts highly suspicious transactions on hold and requests permit of the responsible supervisor].*

Additional Information for Study 4.2b

Exclusion of Participants

In total, the survey was started 220 times. As pre-registered, we excluded participants without management experience (19 cases), working less than 19 hours a week (9 cases), failing at least one attention check (6 cases). All of the remaining participants changed the default value for at least one of the acceptance measures and were fluent in English. Thus, the eligible sample consisted of $N = 186$ participants (as reported in the manuscript).

Pre-Registration

Created:	01/10/2023 04:53 AM (PT)
Made public:	Not yet, will be made publicly available after publication
Available at:	https://aspredicted.org/CSK_3J7 (anonymized, for peer review)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

H1: Usage of AI in companies is more accepted in the area of finances (vs the area of human resources).

H2: The effect of AI functionality on acceptance is dependent on the application area.

H2a: In human resources, usage of AI is less accepted the higher the functionality.

H2b: In finances, usage of AI is more accepted the higher the functionality. However, when functionality reaches a certain level (i.e., implementing), acceptance declines.

3) Describe the key dependent variable(s) specifying how they will be measured.

Acceptance:

- *Risk of negative consequences (1 item)*
- *Willingness to invest (1 item)*

4) How many and which conditions will participants be assigned to?

We will implement a two-factorial mixed-measures design. Participants will be randomly assigned to one potential application area of AI in the business context (between-factor: finances vs human resources).

Each participant will then be exposed to four consecutive subtasks reflecting increasing functionality of the AI (within-factor: 1. Monitoring, 2. Generating, 3. Selecting, 4. Implementing). The order of the subtasks will not be randomized.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

For each of the two acceptance measures (see point 3), we will calculate a mixed ANOVA with application area (finances vs human resources) as between-factor, functionality as within-factor, and the respective acceptance measure as dependent variable. In case of an interaction between application area and functionality we will conduct separate mixed ANOVAs for the two application areas using contrast-coding as follows for functionality (monitoring, generating, selecting, implementing): [c1] -1/4, -1/4, -1/4, 3/4, [c2] -2/3, 1/3, 1/3, 0; [c3] 0, -1/2, 1/2, 0.

Evidence for H1 will be provided by a main effect of application area, whereas evidence for H2 will be provided by an interaction between application area and functionality. Evidence for H2a and H2b will be provided by significant results for the respective contrasts in the follow-up analyses.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Inclusion requirements for participants:

- *Experience being in a management position*
- *Working at least 19 hours a week*
- *Passing two attention checks and stating two have answered all questions attentively.*

We will further exclude participants who do not change the default value (i.e., 0) for any of the acceptance measures (i.e., participants with a 0-value on all four measures of acceptance for all four functionalities will be excluded).

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will use sequential testing: In case the first data collection ($N = 200$) reveals an effect of $f \geq 0.15$ (non-significant) for the main effect of application area or the interaction in the mixed ANOVAs described under (5), we will continue data collection. Based on the observed effect size at this point, we will determine the sample size of the second data collection considering alpha-error cumulation following the recommendations of Lakens (2014). In case both effects are $f < 0.15$ we will not collect further data.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will further assess:

- *Perceived competence of AI for each task (1 item per task)*
- *Perceived integrity of employees for each task (1 item per task)*
- *Reasons why AI usage could be beneficial / risky (2 open-ended questions per task)*
- *Perceived potential to save working hours (1 item) in the respective application area (i.e., finances or human resources).*

Deviations from Pre-Registration

We slightly adjusted the wording of the preregistration in the paper to be consistent across studies (see Table D5 for all changes in the wording of Study 4.2b).

Table D5

Changes in Wording of Study 4.2b

Change	Wording in preregistration	Wording in paper
1	application area	business area
H1	Usage of AI in companies is more acceptable in the area of finances (vs the area of human resources).	Usage of AI is less accepted in HR than in other business areas (i.e., finances, marketing).
RQ	H2a: In human resources, usage of AI is less accepted the higher the functionality. H2b: In finances, usage of AI is more accepted the higher the functionality. However, when functionality reaches a certain level (i.e., implementing), acceptance declines.	Does functionality influence AI acceptance and does this effect depend on the business area?

Materials of Study 4.2b

Procedure of Study 4.2b

1. Informed consent and prescreening
2. Current use of AI in company (1 item)
3. Potential to save working hours in HR/Finances (depending on experimental condition)
4. Description of AI functionalities in HR/Finances (depending on experimental condition, see below)
5. **Willingness to invest, perceived risk of negative consequences** (1 item each per functionality, see manuscript), perceived competence of AI, perceived integrity of

employees, reasons why AI could be beneficial, reasons why AI could risky (1 item each per functionality)

6. Attention check (1 item)
7. Demographics, final consent, option for comment

Description of AI functionalities in HR [Finances]

Study materials were identical to Study 4.2b with changes in HR selection and implementation as reported below.

HR Selection:

Based on the ratings of work-related social media profiles, the AI selects certain people that should be considered in the further recruitment process.

HR Implementation:

The AI autonomously decides about the consideration of persons in the further recruitment process. In detail, the AI decides whom to invite to a short online assessment that serves as the first step in the personnel selection process.

Additional Information for Study 4.3

Exclusion of Participants

The survey was sent out to more than 10,000 companies and we received responses from 1,455 companies. As pre-registered, we excluded data from companies with less than five employees (146 cases), All remaining companies belonged to the knowledge-intensive sector or the manufacturing industry. We further excluded 89 cases with missing data for the dependent variables (i.e., willingness to use in HR and marketing) of our main analysis. Thus, the eligible sample consisted of N = 1,220 as reported in the mansucript.

Pre-Registration

Created:	03/30/2023 07:39 AM (PT)
Made public:	Not yet, will be made publicly available after publication
Available at:	https://aspredicted.org/8SL_WYL (anonymized, for peer review)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

H1: Usage of AI in companies is more accepted in the sector of marketing (vs the sector of human resources).

Based on our data from previous studies we further expect:

H2a: Usage of AI in companies is more accepted the higher the functionality (implementation > generation).

However, we initially assumed the opposite effect:

H2b: Usage of AI in companies is less accepted the higher the functionality (implementation < generation).

To ensure that our findings from prior studies are not based on specific aspects of the chosen methodology, we altered the design (= exchanged between- and within-subjects factors) and the material (= chose a stronger manipulation for the implementation functionality in human resources) in the present study.

RQ: Is the effect of AI functionality on acceptance dependent on the business sector?

3) Describe the key dependent variable(s) specifying how they will be measured.

Willingness to use (1 item per business sector)

4) How many and which conditions will participants be assigned to?

We will implement a two-factorial mixed measures-design. Participants will be randomly assigned to one AI functionality (between-factor: generation vs implementation). Each participant will read two application scenarios of AI with the respective functionality in different business sectors (within-factor: human resources vs marketing). The order of the business sectors will not be randomized.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will calculate a mixed ANOVA with business sector (human resources vs marketing) as within-factor, functionality (generation vs implementation) as between-factor, and willingness to use as dependent variable. Evidence for H1 will be provided by a main effect of business sector, whereas evidence for H2a/b will be provided by a respective main effect of functionality. Evidence for RQ will be provided by an interaction between business sector and functionality.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

The survey will be distributed to a selected target group (see point 7). Companies who do not match the criteria described under (7) will be excluded. We will include participants for analysis if they provide answers for all variables included in the respective analysis.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will recruit participants via a quarterly survey distributed regularly to German companies with at least 5 employees in the knowledge-intensive service sector as well as the manufacturing industry. The survey is sent out to more than 10,000 companies. In the last wave, approximately 1,800 companies answered the survey. For this wave, we expect a somewhat lower response rate. Companies will receive a reminder for participation after 2 weeks, data collection will be stopped after approximately 4-5 weeks.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will further assess:

- *Current Usage of AI in company (2 items)*
- *Perceived quality of current AI tools (2 items)*
- *Required AI quality for usage (1 item per business sector)*
- *Relevance of task (1 item per business sector)*

The data collected for this study is part of a larger survey.

Deviations from Pre-Registration

We slightly adjusted the wording of the preregistration in the paper to be consistent across studies (see Table D6 for all changes in the wording of Study 4.3).

Table D6

Changes in Wording of Study 4.3

Change	Wording in preregistration	Wording in paper
1	Business sector	Business area
H1	Usage of AI in companies is more acceptable in the sector of marketing (vs the sector of human resources).	Usage of AI is less accepted in HR than in other business areas (i.e., finances, marketing).
RQ	H2a: Usage of AI in companies is more accepted the higher the functionality (implementation > generation). H2b: Usage of AI in companies is less accepted the higher the functionality (implementation < generation). RQ: Is the effect of AI functionality on acceptance dependent on the business sector?	Does functionality influence AI acceptance and does this effect depend on the business area?

Materials of Study 4.3

In addition to the scales reported in the manuscript, we further measured current usage of AI in company (2 items), perceived quality of current AI tools (2 items), required AI quality

for usage (1 item per business area), relevance of task (1 item per application area). The data collected for this study is part of a larger survey. German study materials and English translations are reported below.

Description of AI functionalities in HR [Finances]

Generation: Human Resources (German)

Im Personalbereich kann KI eingesetzt werden, um den Rekrutierungsprozess zu optimieren. Dazu kann KI in Jobportalen und Online-Netzwerken (z.B. Xing, StepStone) Profile scannen und hinsichtlich ihrer Passung zu freien Stellen bewerten. Dem Personalbereich stehen im Anschluss die Bewertungen aller gescannten Profile zur Verfügung, um eine eigene Auswahl an Personen zu treffen, die zu einem Bewerbungsgespräch eingeladen werden.

Generation: Human Resources (Translation)

In HR, AI can be used to optimize the recruitment process. For this purpose, AI can scan profiles in job portals and online networks (e.g., Xing, StepStone) and evaluate them in terms of their fit with vacancies. The HR department then has the ratings of all scanned profiles at its disposal to make its own selection of people to invite for an interview.

Generation: Marketing (German)

Im Marketing kann KI eingesetzt werden, um eigene Werbekampagnen zu optimieren. Dazu kann KI Daten vorheriger Werbekampagnen analysieren und mögliche Werbeinhalte für neue Kampagnen (z.B. Social Media Posts) generieren. Dem Marketingbereich steht im Anschluss eine Auswahl KI-generierter Inhalte zur Verfügung, um zu entscheiden, welche Inhalte veröffentlicht werden (z.B. auf Social-Media-Kanälen oder der Webseite des Unternehmens).

Generation: Marketing (Translation)

In marketing, AI can be used to optimize own advertising campaigns. For this purpose, AI can analyze data from previous advertising campaigns and generate possible advertising content for new campaigns (e.g., social media posts). The marketing department then has a selection of AI-generated content at its disposal to decide which content to publish (e.g., on social media channels or the company's website).

Implementation: Human Resources (German)

Im Personalbereich kann KI eingesetzt werden, um den Rekrutierungsprozess zu optimieren. Dazu kann KI in Jobportalen und Online-Netzwerken (z.B. Xing, StepStone)

Profile scannen, hinsichtlich ihrer Passung zu freien Stellen bewerten und eine Auswahl geeigneter Personen treffen. Die von der KI ausgewählten Personen werden automatisch und ohne eine Prüfung durch den Personalbereich zu einem Bewerbungsgespräch eingeladen.

Implementation: Human Resources (Translation)

In HR, AI can be used to optimize the recruitment process. For this purpose, AI can scan profiles in job portals and online networks (e.g., Xing, StepStone), evaluate them in terms of their fit with vacancies, and make a selection of suitable candidates. The people selected by the AI are automatically invited to an interview without a check by HR.

Implementation: Marketing (German)

Im Marketing kann KI eingesetzt werden, um eigene Werbekampagnen zu optimieren. Dazu kann KI Daten vorheriger Werbekampagnen analysieren und mögliche Werbeinhalte für neue Kampagnen (z.B. Social Media Posts) generieren. Die von der KI generierten Inhalte werden im Anschluss automatisch und ohne eine Prüfung durch den Marketingbereich veröffentlicht (z.B. auf Social-Media-Kanälen oder der Webseite des Unternehmens).

Implementation: Marketing (Translation)

In marketing, AI can be used to optimize own advertising campaigns. For this purpose, AI can analyze data from previous advertising campaigns and generate possible advertising content for new campaigns (e.g., social media posts). The content generated by the AI is then automatically published (e.g., on social media channels or the company's website) without any review by the marketing department.

Summary

Humans interact more and more frequently with evermore capable AI (Artificial Intelligence) systems. While these interactions offer many opportunities, they also come with certain risks. In cases in which AI is designed to give personalized output, this becomes particularly evident: on the one hand, people can benefit from personalized interactions (e.g., receiving personalized recommendations or a personalized usage experience in using conversational AI), but on the other hand, personalization requires the collection of personal data which is often accompanied by concerns for users' privacy. As people nonetheless often use such AI systems and willingly disclose their personal data, the current dissertation aimed at a better understanding of disclosure in the increasingly common context of human-AI interactions.

For this purpose, this dissertation used a human-oriented approach building on the ideas of the computers as social actors paradigm (Nass & Moon, 2000; Reeves & Nass, 1996) and anthropomorphism (Epley et al., 2007). Three empirical chapters investigated how perceptions of (1) the interaction partner (i.e., the technology and its provider) and (2) the interaction (i.e., perceived interactivity and output quality) are related to individuals' decisions to disclose personal information in private use contexts of AI. Furthermore, considering that individuals cannot always decide whether AI is used – for instance, in the work context – this dissertation addressed the perspective of decision-makers (i.e., managers) and how their AI acceptance is associated with perceptions of AI and the relevance of personal data in certain business areas.

Taken together, the findings of the current dissertation highlight that the human-oriented approach (i.e., focusing on how humans perceive the interaction partner and the interaction rather than the technical details of implementation) is useful to get a better understanding of disclosure in human-AI interactions. Within several empirical studies, the current thesis showed that perceptions of the interaction partner (i.e., the technology and the provider) as well as of the interaction (i.e., regarding the perceived interactivity and output quality) seem important for individuals' disclosure (and decision-makers' AI acceptance). Most importantly, it became evident that the capabilities of the AI as interaction partner can evoke positive reactions (e.g., users were more willing to disclose information if perceiving intellectual competencies in a conversational AI), but – if they cross a critical boundary – also lead to adverse reactions (e.g., non-users being less willing to disclose and users and non-users being more concerned about privacy if perceiving meta-cognitive heuristics in conversational AI; managers reporting lower AI acceptance for implementation). Thus, it seems that very high capabilities of AI might not always be perceived as desirable. Accordingly, future research needs to grasp a better understanding of why such negative reactions arise and how we can ensure a safe interaction of

humans with AI technologies that harness the full potential of these increasingly capable technologies while avoiding – or at least minimizing – the associated risks and reducing adverse reactions.

Deutsche Zusammenfassung

Menschen interagieren immer häufiger mit zunehmend leistungsfähiger KI (Künstliche Intelligenz) bzw. KI-basierten Systemen. Diese Interaktionen bieten zwar viele Chancen, bergen zugleich aber auch gewisse Risiken. In Fällen, in denen KI personalisierte Ergebnisse liefern soll, wird dies besonders deutlich: Einerseits können Menschen von personalisierten Interaktionen profitieren (z. B. durch personalisierte Empfehlungen oder ein personalisiertes Nutzungserlebnis bei der Verwendung von KI-basierten Sprachassistenten). Andererseits müssen hierfür persönliche Daten der Nutzenden gesammelt, gespeichert und ausgewertet werden, was häufig zu Datenschutzbedenken führt. Da entsprechende KI-Systeme dennoch viel genutzt werden und Menschen bereitwillig ihre persönlichen Daten preisgeben, ist das Ziel der vorliegenden Dissertation ein besseres Verständnis davon zu erlangen, wann Menschen bereit sind in Interaktionen mit KI-Systemen ihre Daten zu teilen (und wann sie sich um ihre Privatsphäre sorgen).

Hierzu wurde in der vorliegenden Arbeit ein menschenorientierter Ansatz verwendet, der auf den Ideen des Computer-als-soziale-Akteure-Paradigmas (Nass & Moon, 2000; Reeves & Nass, 1996) und des Anthropomorphismus (Epley et al., 2007) aufbaut. In drei empirischen Kapiteln wurde untersucht, wie die Wahrnehmung (1) des Interaktionspartners (d.h. der Technologie und ihres Anbieters) und (2) der Interaktion (d.h. der wahrgenommenen Interaktivität und der Ergebnisqualität) damit zusammenhängen, ob Individuen persönliche Informationen mit KI-Systemen teilen, die sie privat nutzen. In Anbetracht der Tatsache, dass Individuen nicht immer selbst entscheiden können, ob KI genutzt wird – beispielsweise im Arbeitskontext – wurde in dieser Dissertation außerdem die Perspektive von Entscheidungsträgern (Managern) beleuchtet und untersucht, inwiefern deren KI-Akzeptanz mit der Wahrnehmung der KI als Interaktionspartner und der Relevanz persönlicher Daten in verschiedenen Geschäftsbereichen zusammenhängt.

Zusammengefasst zeigen die Ergebnisse, dass der menschenorientierte Ansatz (mit Fokus auf die menschliche Wahrnehmung des Interaktionspartners und der Interaktion statt auf Details der technischen Implementierung) nützlich ist, um besser zu verstehen, wann Menschen bereit sind Daten mit KI-Systemen zu teilen. Die durchgeführten empirischen Studien weisen darauf hin, dass für die Entscheidung, ob persönliche Daten geteilt werden, die Wahrnehmung des Interaktionspartners (d.h. der Technologie und des Anbieters) sowie der Interaktion (d.h. der wahrgenommenen Interaktivität und der Ergebnisqualität) eine wichtige Rolle spielen. Es wurde insbesondere deutlich, dass wahrgenommene Fähigkeiten der KI als Interaktionspartner sowohl positive Reaktionen hervorrufen können (z. B. waren Nutzer eher bereit, Informationen

preiszugeben, wenn sie intellektuelle Kompetenzen in einer KI wahrnahmen), aber – wenn die KI-Fähigkeiten eine kritische Grenze überschreiten – auch zu negativen Reaktionen führen können (z. B. waren Nicht-Nutzer weniger bereit, Informationen preiszugeben, wenn sie metakognitive Heuristiken in einer KI wahrnahmen; Manager zeigten eine geringere Akzeptanz von KI, die eigenständig implementiert). Es ist daher naheliegend, dass sehr hohe Fähigkeiten von KI nicht immer als positiv wahrgenommen werden. Daher sollte künftige Forschung adressieren, warum solche negativen Reaktionen gegenüber KI auftreten und wie wir eine sichere Interaktion zwischen Menschen und KI-Technologien gewährleisten können, die das volle Potenzial dieser immer leistungsfähigeren Technologie ausschöpft und gleichzeitig die damit verbundenen Risiken vermeidet – oder zumindest minimiert – und negative Reaktionen reduziert.

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die zur Promotion eingereichte Arbeit mit dem Titel „Towards a Better Understanding of Information Disclosure in Human-AI Interactions“ selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich erkläre, dass die Richtlinien zur Sicherung guter wissenschaftlicher Praxis der Universität Tübingen (Beschluss des Senats vom 25.5.2000) beachtet wurden. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

Tübingen, _____ _____

Unterschrift