6-18-2022

WHAT SHOULD AI KNOW? INFORMATION DISCLOSURE IN HUMAN-AI COLLABORATION

Izabel Cvetkovic
University of Hamburg, izabel.cvetkovic@uni-hamburg.de

Sarah Oeste-Reiß
University of Kassel, oeste-reiss@uni-kassel.de

Nale Lehmann-Willenbrock
University of Hamburg, nale.lehmann-willenbrock@uni-hamburg.de

Eva A. C. Bittner
University of Hamburg, bittner@informatik.uni-hamburg.de

WHAT SHOULD AI KNOW? INFORMATION DISCLOSURE IN HUMAN-AI COLLABORATION

Research in Progress

Izabel Cvetkovic, Universität Hamburg, Hamburg, Germany, izabel.cvetkovic@uni-hamburg.de
Sarah Oeste-Reiß, Universität Kassel, Kassel, Germany, oeste-reiss@uni-kassel.de
Nale Lehmann-Willenbrock, Universität Hamburg, Hamburg, Germany,
nale.lehmann-willenbrock@uni-hamburg.de
Eva Bittner, Universität Hamburg, Hamburg, Germany, eva.bittner@uni-hamburg.de

Abstract

AI-assisted Design Thinking shows great potential for supporting collaborative creative work. To foster creative thinking processes within teams with individualized suggestions, AI has to rely on data provided by the teams. As a prerequisite, team members need to weigh their disclosure preferences against the potential benefits of AI when disclosing information. To shed light on these decisions, we identify relevant information such as emotional states or discussion arguments that design thinking teams could provide to AI to enjoy the benefits of its support. Using the privacy calculus as theoretical lens, we draft a research design to analyze user preferences for disclosing different information relevant to the service bundles that AI provides for respective information. We make explorative contributions to the body of knowledge in terms of AI use and its corresponding information disclosure. The findings are relevant for practice as they guide the design of AI that fosters information disclosure.

Keywords: Human-AI Collaboration, Information Disclosure, Design Thinking, UI Design.

1 Introduction

To remain competitive in the fast-changing technology landscape, characterized by digitization and globalization, companies must rely on novel ideas to develop innovative and profitable products, processes, and services. A promising approach in this context is Design thinking (DT) (Meinel and Leifer, 2012). It is characterized as a problem-solving approach, based on collaboration, user-centered design, and creativity (Schallmo, 2017). However, DT can incur challenges in terms of tasks, team interaction, and method-specific facilitation (Bittner et al., 2021), such as an inappropriate division of tasks, ineffective selection of methods, and destructive workshop dynamics (Seeber et al., 2020). Fast advances in Artificial Intelligence (AI) provide promising potential to cope with these challenges. AI is the science and engineering behind creating intelligent machines, particularly computer programs, which try to grasp and, to some extent, imitate human intelligence (McCarthy, 2007). Recent research outlines the possibility of supporting DT approaches with AI, for instance, by means of virtual collaborators (Siemon and Strohmann, 2021) that support human teams and facilitators in specific tasks, such as persona building (Lembcke et al., 2020). This indicates that Human-AI Collaboration in DT is a promising research avenue.
Human-AI Collaboration, i.e., Hybrid Intelligence (Dellermann et al., 2019b) is a flourishing and important research area (Bittner et al., 2019b; Dellermann et al., 2019a; Seeber et al., 2020). It combines the complementary strengths of humans and AI to achieve better performance than humans or AI could

achieve alone. Some examples of such hybrid systems can be seen in decision making practices, where AI recommends actions to human experts who are responsible for final decisions (Bansal et al., 2021). In the course of this, a benefit of AI stems from analyzing large amounts of data and giving recommendations based on insights resulting from these analyses (Cranor, 2008). The human verifies AI recommendations, and either chooses the most appropriate solution or gives feedback to the system, from which it can consequently learn and improve upon. Such feedback can be used to retrain the models to be reused in the future, or the models can be updated in real-time and output improved accordingly (Holzinger, 2016).
In any case, AI learns and improves based on data (Russell and Norvig, 1995). In Human-AI collaborative scenarios, AI can adopt the role of a teammate (Seeber et al., 2020) or a co-facilitator (Bittner and Shoury, 2019). For example, in DT approaches it would make more sense if AI made use of the data being collected during the session and present it back in a structured manner, thereby promoting creative work. AI could collect different types of information, such as emotional and cognitive states or negative utterances to moderate team dynamics. Specifically, AI could use social signals from the team to capture bad dynamics and prevent conflicts (Elshan et al., 2022).
However, problems can arise in regards to the disclosure of social signals as input for AI. Experience has shown that AI-based detection of social signals such as emotion recognition are followed by low disclosure of such information (Bailenson et al., 2006). Since the nature of this technology is psychologically invasive, privacy issues must be dealt with to the satisfaction of all persons involved in the system. On the other hand, an affect recognition system could present group members with relevant information about the group's general emotional state. Members of a group, or a facilitator, could use this information to assess the strength of agreement or disagreement for specific positions (Salim and Yusoff, 2008). State-of-the-art products in this domain demonstrate how social signals can be leveraged to receive benefits in return. For example, MIT's social robot TEGA tracks children's emotions and uses this information to offer better learning assistance (Westlund et al., 2016). Therefore, AI service bundles should come with explicit benefits that human users will receive when disclosing information. Accordingly, exploratory research on user preferences and quid pro quos between disclosure risks and benefits is needed.
This work aims to identify the service bundles that an AI-based system can offer a human user in exchange for the user's information disclosure (ID) and gather insights toward such input-output service pairs. Service bundles are combined services that offer a reward benefit for a user (Andrews et al., 2010). Important factors in human-centred AI development are human needs and preferences, as these motivate human users toward using a system. An important avenue hereby is ID (Rayo and Segal, 2010), as information creates the foundation for system recommendations (and lifelong learning systems). However, to date, little is known about human preferences to disclose information in collaborative scenarios with AI-based systems, i.e., scenarios from which they have direct, visible profit in terms of task assistance, better and faster performance, or better-quality results.
To address this research gap, this work pursues to answer the following research questions: First, we need to identify the information that humans can disclose in our specific scenario that can subsequently be used by AI. Hence, we pose the RQ 1: *Which provided input from humans can be processed by AI-based systems to support them in specific tasks in creative teamwork?* Inputs identified by addressing RQ 1 are different regarding the type of information they carry (e.g., emotional states or factual information), and the relevance this type might have to disclosure preferences. Therefore, we pose the RQ 2: *Which information and to what extent are humans willing to disclose so that AI-based systems can support them in specific tasks in creative teamwork?* Finally, in order to investigate the effect of AI service bundles to ID, we pose the RQ 3: *How should the output from AI-based systems be designed to foster information disclosure in specific tasks in creative teamwork?*

2 Theoretical Background

2.1 Design Thinking and Artificial Intelligence

DT is an approach that integrates various tools and techniques for collaborative problem solving (Liedtka, 2015). This method provides an appropriate use case scenario for research on Human-AI collaboration due to its characteristics (Schallmo, 2017), and its increasing relevance for creative knowledge work, such as innovation development (Meinel and Leifer, 2012). In DT approaches, a human facilitator guides the collaboration process, thereby moderating the team dynamics and fostering creativity inter alia (Bittner et al., 2021). DT is used in various industries such as software, healthcare, automotive, retail, and finance. Companies use DT approaches for a range of purposes such as more empathetic customer observations, concept generation, prototyping, and product development (Carlgren et al., 2014). A common DT process comprises five stages, namely empathizing, defining, ideating, prototyping, and testing (Brown and Katz, 2011).

These are moderated throughout by human facilitators. Facilitators are responsible for the appropriate support of the teams and their challenges, as well as for the management and organization of the entire DT process: measures such as team composition and stakeholder management, process and method planning, provision of materials, documentation, and overview and transfer to the implementation of the developed ideas are tasks that require a lot of time and organization (Hjalmarsson et al., 2015). Due to the mentioned characteristics of the DT method and the high workload on facilitators, this approach could particularly profit from AI assistance (Bittner et al., 2021). Debowski et al. (2021) outline the problem areas in creativity workshops and present design principles for a virtual assistant to address these problems. Strohmann et al. (2018) demonstrate the need for AI assistance in DT and outline the requirements for such assistance in terms of conditions, characteristics, and tasks. Similarly, Bittner et al. (2021) outline the need for AI assistance in different DT areas, namely: Team Task and Process Facilitation, Team Interaction Facilitation, Facilitator Task and Process Assistance, Facilitator Interaction Assistance, and Method Specific Assistance.

Human facilitators use various types of human input to moderate DT, such as emotional and cognitive states, arguments from discussions, ideas etc. However, whenever AI is involved, the issues of ethics and data privacy arise (Manheim and Kaplan, 2018). These issues are very context-specific, i.e., some data is more sensitive, such as user behavior data, and some is more easily disclosed, e.g., demographics (Knijnenburg and Kobsa, 2013a). AI-assisted DT is in an early phase of research (Debowski et al., 2021). Research in other domains has shown that people tend to accept new technologies without worrying much about privacy if they get sufficient benefits from them (Salim and Yusoff, 2008). Furthermore, there is overwhelming evidence that people will trade personal data in exchange for services or social benefits (Pitt, 2012; Youn, 2009). However, ID also comes with perceived risks (Al-Natour et al., 2021). Notwithstanding, little is known about how people would behave with regards to ID in DT, i.e., whether they would be comfortable providing all the data to AI that they usually provide to a human facilitator. Nevertheless, the relevance of information exchange in terms of feedback and direct communication for DT is undeniable. In the proposed work, we want to tackle the team interaction space of DT, since this space most likely requires sensitive data from participants for successful teamwork facilitation.

2.2 Information Disclosure

ID refers to the extent to which users are willing to disclose certain private information, which is often perceived as an issue (Knijnenburg and Kobsa, 2013a). Giving users explicit control over the information they disclose is one way to address this issue. In this case, ID becomes a decision, which comes down to calculating the benefits and risks of making a certain decision (Knijnenburg and Kobsa, 2013b). This concept is also known as "privacy calculus" (Laufer and Wolfe, 1977). A common example of such a cost-benefit trade-off is ID on social media. Here, users gain social benefits in exchange for information

but are also exposed to loss of privacy. The problem of privacy calculus is the intangibility of the value of privacy, its context-dependence, and uncertainty regarding privacy decisions. The latter is due to not knowing how the data will be used, i.e., missing transparency (Hirschprung et al., 2016). ID thus depends on situational and context-specific factors that directly affect the privacy calculus (Xu et al., 2008).
ID has been studied in a number of contexts such as e-commerce (Berendt et al., 2005), or social networks (Kroll and Stieglitz, 2021) through several theoretical lenses. An integrated framework of theories in privacy research (Li, 2012) lists theories most often related to this research such as social exchange theory (Thibaut and Kelley, 1959), social contract theory (Milne and Gordon, 1993), theory of reasoned action (Ajzen and Fishbein, 1980), and privacy calculus theory (Laufer and Wolfe, 1977). For example, social exchange theory posits that people decide whether to engage in a social act based on an evaluation of interpersonal rewards and costs associated with the act. Applied to information exchange this means that when perceived benefit is higher than perceived risk, ID is more likely to happen (Chang and Heo, 2014; Li et al., 2010). Social exchange also involves trust, and cannot be reduced to simple bargaining (Masaviru, 2016). Common to most of the mentioned theories is the aforementioned cost-benefit trade-off, i.e., the privacy calculus. Hence, we choose to adopt the privacy calculus theory as our kernel theoretical lens.
ID can be affected by rewards, order of disclosure requests, justifications, and privacy indicators or statements. For example, justifications can include reasons for requesting information, benefits of disclosure, or appealing to social norms, i.e., showing that others decided to disclose the same information (Knijnenburg and Kobsa, 2013b). Privacy statements such as the TRUSTe seal increase the likeliness of disclosing information (Xu et al., 2009). Furthermore, increased system transparency has a positive effect on the disclosure of information (Patil and Kobsa, 2005; Patil and Lai, 2005). Users who have more choice over how, when and with whom components of their environment are shared can establish a better balance between awareness and privacy (Patil and Lai, 2005). Apart from privacy self-management, transparent privacy policies and privacy controls can further alleviate concerns about disclosure (Stutzman et al., 2011).
A way to encourage ID without restricting users' freedom of choice is digital nudging (Kroll and Stieglitz, 2021). Digital nudging represents the use of user-interface design elements to shape people's behavior in digital environments (Weinmann et al., 2016). For example, Facebook's privacy dinosaur helps navigate complex privacy options and make choices with less overload (Kroll and Stieglitz, 2021).
The introduction of AI and its relation to data collection raises privacy concerns and increases cautious behavior. Reliable AI assistance depends on user data. For example, Netflix depends on users' viewing history and movie ratings to provide suitable recommendations (Liao and Sundar, 2021). Usually, such services also offer an explanation that emphasizes the benefit of data disclosure. Intelligent personal assistants represent another form of AI often related to privacy concerns. For example, Manikonda et al. (2018) found that users who mute their microphones state privacy concerns as the main reason for this. An empirical study of ID to virtual advisors investigated a number of determinants of the disclosure, such as trust, transparency, and perceived benefits and costs. The authors found that all determinants contained in their model significantly contribute to ID (Al-Natour et al., 2021). When it comes to ID to AI in the workplace, trust in the company plays another important role for weighing the risks and benefits of disclosure (Metzger, 2004). A way to influence the determinants of disclosure to AI is by using protection frameworks such as federated learning, which use decentralized data sharing techniques to perserve information privacy (Wei et al., 2020).
An example of sensitive data collection in the workplace that has been discontinued due to privacy reasons is Zoom's attention tracker (*Attendee attention tracking*). Such examples call for more research in the area of privacy and motivate the question: How can we reconcile the need for massive private data collection for AI and privacy concerns? The proposed study tries to tackle an aspect of this question by gaining insight into preferences for disclosure relative to benefits received in exchange.

3 Preliminary Results and Proposed Research Approach

DT processes could profit from AI assistance in different ways. We are analyzing literature to explore this in-depth and identify the input/output space of Human-AI collaboration in DT. The main paper, from which we extend the literature analysis is "Digital Facilitation Assistance for Collaborative, Creative Design Processes" (Bittner et al., 2021).

	Human Input	**AI Output**
Team interaction	Body language (gestures, facial expressions) (Benke et al., 2020; Bittner et al., 2021)	Team mood board (own suggestion); jokes and anecdotes in response to negative mood (Strohmann et al., 2018)
	Body language (eye gaze, mimics, gestures) (Benke et al., 2020; Bittner et al., 2021)	Reminder to take a break, information load board (own suggestion)
	Keywords for breaking rules (Bittner et al., 2021)	Reminder of the rules of creative teamwork (Bittner et al., 2021)
	Critical or negative utterances or behavior (Bittner et al., 2021; Starostka et al., 2021)	Reminder to keep the DT mindset; Explanation on how to express constructive feedback (Bittner et al., 2021)
	Speech share & centrality (Bittner et al., 2021; Przybilla et al., 2019)	Informing about unbalanced speech shares to promote equal participation (Leimeister, 2014)
Team process	Results from a brainstorming session (Bittner et al., 2021)	Structured and clustered results presented in a dashboard (Bittner et al., 2021)
	Calendar information (Cranshaw et al., 2017)	Recommended schedule for subsequent session(s) (own suggestion)
	Transcripts of the workshop (Bittner et al., 2021)	Previous session(s) recap (own suggestion)
	Data from interviews/desk research (Bittner et al., 2021)	Analysis report (e.g., Recommendation for core categories from interviews) (Bittner et al., 2021)
	Stagnation (e.g., not enough ideas within 10 minutes of brainstorming) (Seeber, 2019; Strohmann et al., 2018)	Proactive acting - stimulus topic or motivational quote (Strohmann et al., 2018)

Table 1: Examples of DT input identified in the literature and proposed AI output for each respective input

An AI-based system could support team processes by e.g., summarizing and clustering results from user, market, and problem analysis and presenting them in a more comprehensible way (Bittner et al., 2021; Lembcke et al., 2020). Furthermore, AI could assist with meeting planning and organization by retrieving calendar information (Cranshaw et al., 2017) from DT participants and suggesting suitable time slots for the subsequent sessions, or it could use transcriptions from previous sessions (Bittner et al., 2021) to create recaps during subsequent sessions.

Regarding the DT area of Team Interaction, we found even more cases where AI could meaningfully assist. For instance, by using different inputs from body language (mimics, gestures, eye gaze, facial expressions), AI could create a team mood board, based on which it could step in to lighten the mood with jokes or anecdotes, in case the mood is shifting to negative (Benke et al., 2020; Bittner et al., 2021; Strohmann et al., 2018). By analyzing eye-movement data, AI could create an information load board and suggest a break at a suitable time point (Fig. 1b). Furthermore, based on these input data, AI could identify personal characteristics of the participants in order to provide more tailored interventions when a critical situation occurs (Strohmann et al., 2018), or it could act as an Animator or Game Master that motivates the team by inducing mood-lightening recreational interventions such as games or jokes when participants get exhausted in the course of a long workshop (Bittner et al., 2021). Moreover, to foster team interaction, AI could infer team states from extracted human signals such as speech shares (Fig. 1a) and centrality (Bittner et al., 2021; Clawson et al., 1993; Leimeister, 2014; Strohmann et al., 2018), critical or negative utterances or behavior (Bittner et al., 2021), and even a lack thereof, in which case it could act proactively by providing a stimulus topic or a motivational quote (Strohmann et al., 2018).

Apart from Team Facilitation, AI could also serve as Facilitator Assistance and Method Specific Assistance. In the former case, AI could support both process and interaction facilitation. For example, by analyzing personal profiles of participants, AI could foster team building by employing an unbiased composition of teams, e.g., with sufficient diversity (Bittner et al., 2021; Cautela et al., 2019; Xiao et al., 2019). On the interaction side, AI could use mood indicators to alert a human DT facilitator about negative team dynamics (Bittner et al., 2021; Elshan et al., 2022; Leimeister, 2014).
Examples of Method Specific Assistance can be demonstrated on the Ideation Method. Here, an AI could build upon ideas from participants to support them in various ways, e.g., it could provide individual feedback and inspiration for further ideas (Przybilla et al., 2019; Strohmann et al., 2017), or it could assist by creating an idea database, structuring and rating ideas by similarity, problem to be solved, and innovativeness of the idea (Bittner et al., 2019a, 2021; Ulrich, 2018; Voigt, 2014). Additionally, AI could build upon ideas to create recommendation templates for sketches or other forms of visualization, such as movies or physical modeling through materials such as cardboard (Bittner et al., 2021; Voigt, 2014).

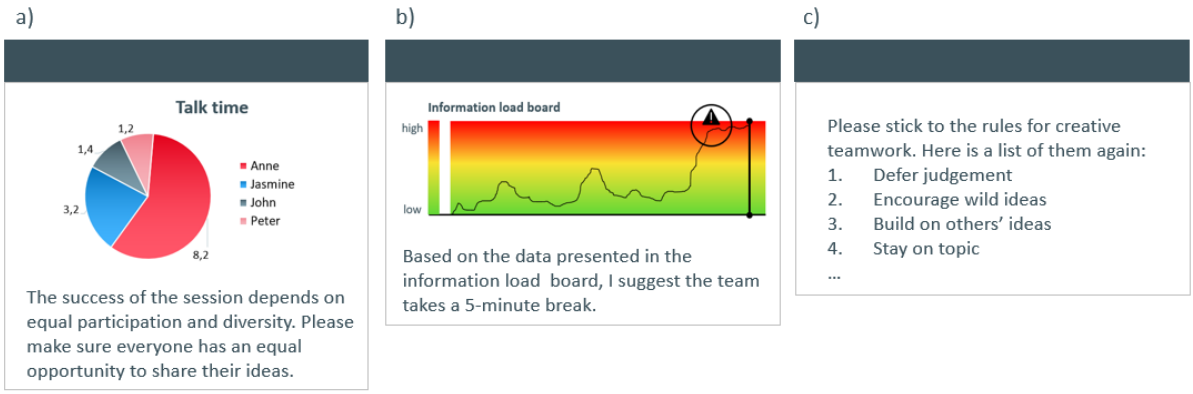

Figure 1: Examples of possible AI outputs for different inputs: a) Unbalanced speech shares. b) Increased cognitive load. c) Breaking rules of creative teamwork.

In light of the findings from the literature, we contend that the AI output design affects the privacy calculus and, thus, ID. Specifically, we hypothesise that H1: *AI Output perceived as beneficial leads to higher willingness to disclose particular information.*; H2: *AI Output perceived as risky regarding loss of privacy leads to lower willingness to disclose particular information.* To address our research questions, an online experiment employing a vignette method will be conducted. To answer the main research question, i.e., RQ 2: "Which information and to what extent are humans willing to disclose so that AI-based systems can support them in specific tasks in creative teamwork?", we will measure ID as an ordinal variable on a 5-point likert scale by asking participants how likely they are to disclose particular information. Independent variables hereby are input content (see examples in Table 1) and for each input the AI output design in terms of recipient, content, and representation. To address our research hypotheses, for each output design perceived benefit and perceived risk of the design will be measured. For each input, we will ask to what extent participants would disclose it, after presenting different AI output designs in terms of recipients and representation. Table 2 provides an overview of the levels of the independent variables.

Input	Output		
Content	Recipient	Content	Representation
Body language	Team	Reminder	Non-transparent
Speech shares	Individual	Dashboard	Transparent anonymized
Negative utterances		Animation	Transparent non-anonymized
Keywords for breaking rules		Explanation	

Table 2: Levels of independent variables. Input and output content is not exhaustive.

We will further analyze the impact of AI output in terms of recipient and representation on ID. To test the hypotheses, we will perform linear regression analyses and corresponding post-hoc tests. Moreover, we will employ the analysis of covariance to test whether different information leads to different disclosure preferences.

The experiment will employ a vignette-based method, which is a proven effective and economical method to assess behaviours, attitudes and intentions. Furthermore, it is a suitable method for exploring sensitive topics, such as ID (Aguinis and Bradley, 2014; Karren and Barringer, 2002). Vignette method employs imagined scenarios, in our case AI is working together with human teams on a task, using the respective inputs as a way to foster task performance and team interaction, as this is the area of potentially meaningful AI assistance in DT with major ID prerequisites. The design of the scenario will be informed via literature review. After presenting each vignette (one vignette corresponding to one AI output design option for one input), participants will be asked to what extent they would disclose certain information to enjoy the benefits of this specific AI assistance. Some vignettes will serve as a reward nudge for ID and will employ principles of transparent AI, such as meaningful explanations that accurately reflect the system's process and emphasize that the system only operates under the conditions for which it was designed (Phillips et al., 2020). These will ensure the delivery of proper explanations for data usage and foster trust. Participants will also be able to select why they would choose not to disclose certain information. An open-end question will be provided to account for factors for non-disclosure not identified in the literature. We will collect further qualitative data regarding participants' suggestions about the possible solutions that could lead to increased disclosure intention.

AI output will vary in terms of content, recipient, and representation. Content can be, e.g., a reminder to follow DT rules (see Fig. 1c), or a joke in response to a negative mood. Participants will choose, whether they would rather accept such output provided to the entire team or only in private. The output representation will vary in terms of transparency and anonymity. Three design options will be presented for each input content and recipient: non-transparent, transparent and anonymized, and transparent and non-anonymized (Fig. 2). The assumption hereby is that different output designs lead to different disclosure preferences in terms of recipients. Due to possible differences in perceived risks and benefits of ID in teams vs. in private, we hypothesize that, e.g., a pie chart with speech shares might rather be received only personally, in an anonymized form, but a suggestion for a break based on a summated information load board might be more beneficial when presented to a whole team. However, we are not locking in assumptions, and aim to derive specific hypotheses for each output in the following stages of this research.

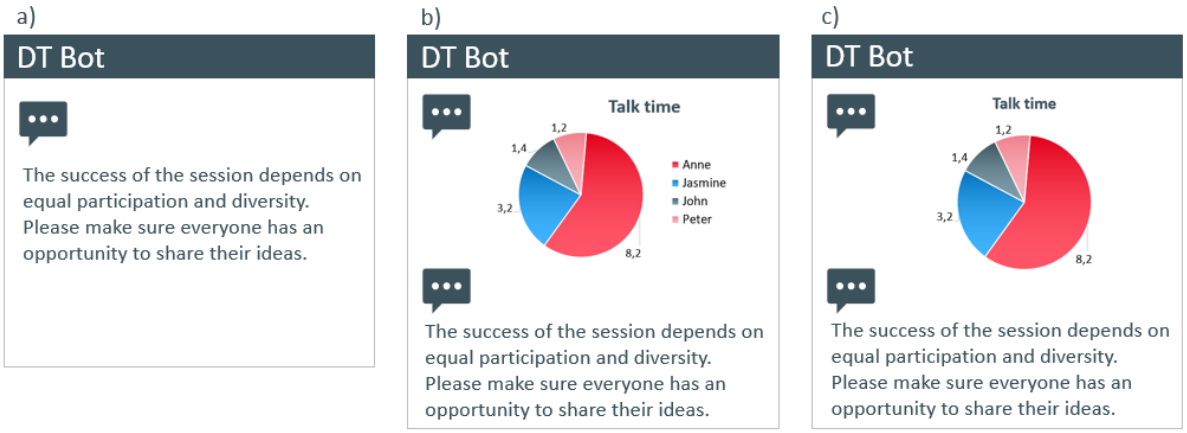

Figure 2: Example of the experimental material. a) Non-transparent output. b) Transparent and non-anonymized output. c) Transparent and anonymized output.

Figure 2 illustrates an example of the experimental material. Here, speech shares of participants are shown by AI, with a hint to keep the speech shares as equal as possible, since this is a very important principle

for successful DT practices. Three design options are considered: in the first one (Fig. 2a) only a hint to equalize talk times is shown. There is no transparency, as participants do not know which input AI used to provide them with this information; in the second variant (Fig. 2b) talk time is shown together with the names of corresponding participants. This is a transparent and non-anonymized representation of output. The pie chart depicts what information (talk time) AI used for this output, thus ensuring transparency. This output is non-anonymized since the chart also displays the names of the participants. The third option (Fig. 2c) only shows that talk time is unequal between participants, without exposing any names. Therefore, this output representation is both transparent and anonymized. The potential design space for AI output is immense, ranging from minimal cues to complex dashboards. We aim to further refine our design space by evaluating results of the proposed study while also continuously supplementing the theoretically guided part by studying current literature.

The experiment will be performed with workers and/or students who have already participated in DT workshops, in order to ensure their understanding of the scenario as well as an understanding of challenges of DT and possible benefits of AI assistance. Due to differences in privacy concerns regarding gender and age (Knijnenburg and Kobsa, 2013a) we aim for a balanced sample, in order to control for these variables in subsequent analyses.

4 Expected Contribution and Conclusion

AI-assisted DT shows great potential for supporting collaborative creative work. Important challenges en route to human-centered AI design for Human-AI collaboration concern ID and ethical data usage. By conducting our study, we plan on contributing to this research stream, in both theoretical and practical terms.

First, the study contributes to the knowledge base by identifying relevant human inputs in DT that could potentially be used by AI to support DT sessions as a teammate or (co-)facilitator. Furthermore, for each identified input, we specify a corresponding AI output, which maximizes the value of the respective input to support teamwork in DT. In doing so, we expect to gain a better understanding of privacy calculus in the given context as well as derive design guidelines for transparent AI that fosters ID and helps to create value in collaborative creative work. By gaining insights about our inital hypotheses we expect to derive more specific hypotheses matched to specific input-output pairs in further iterations. Moreover, we hope to inspire researchers in other areas of AI research to conduct groundwork on AI service bundles for particular inputs, while taking in regard privacy preferences.

Our main findings are expected to contribute to the privacy research in Human-AI collaboration, in terms of information types and respective levels of ID for each type. Beyond that, we expect to gain insights into factors related to disclosure, which we will explore via different AI output designs by directly asking study participants about the reasons for their non-disclosure and further suggestions. The study shall thus identify reasons for non-disclosure, such as trust, risk, and disclosure benefits, and strengthen both theoretical and design knowledge about ID. Additionally, by examining ID in terms of AI output recipient, we expect to shed light on differences and similarities of ID in teams vs. in private. Since many determinants of ID are shared across different use cases, we except to contribute with transferable knowledge by validating these determinants empirically. In addition, perspective development of design patterns shall generalize the findings beyond the specific use case to similar domains.

Limitations of our proposed study include testing only high-level hypotheses within a rather simplistic research model. We plan to address these by drafting a comprehensive research model including various other relevant determinants of ID such as trust and perceived control and analyze it via structural equation modelling. Moreover, we also plan on considering differences in determinants of disclosure in private vs in team and, if necessary, draft separate research models accordingly. Further, just as every method, vignette method comes with its limitations. Finally, measuring intention to disclose comes with a risk of underestimating the privacy paradox, i.e., that in some cases intention to disclose does not correspond to the actual disclosure behavior.

References

Aguinis, H. and K. J. Bradley (2014). "Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies." *Organizational Research Methods* 17 (4), 351–371. DOI: 10.1177/1094428114547952. URL: https://doi.org/10.1177/1094428114547952.

Ajzen, I. and M. Fishbein (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, N.J: Prentice-Hall.

Al-Natour, S., I. Benbasat, and R. Cenfetelli (2021). "Designing Online Virtual Advisors to Encourage Customer Self-disclosure: A Theoretical Model and an Empirical Test." *Journal of Management Information Systems* 38 (3), 798–827. DOI: 10.1080/07421222.2021.1962595. URL: https://doi.org/10.1080/07421222.2021.1962595.

Andrews, M. L., R. L. Benedicktus, and M. K. Brady (2010). "The effect of incentives on customer evaluations of service bundles." *Journal of Business Research* 63 (1), 71–76. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2009.01.011. URL: https://www.sciencedirect.com/science/article/pii/S0148296309000459.

Bailenson, J. N., N. Yee, D. Merget, and R. Schroeder (2006). "The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction." *Presence: Teleoperators and Virtual Environments* 15 (4), 359–372. DOI: 10.1162/pres.15.4.359. URL: https://doi.org/10.1162/pres.15.4.359.

Bansal, G., B. Nushi, E. Kamar, E. Horvitz, and D. S. Weld (2021). "Is the Most Accurate AI the Best Teammate? Optimizing AI for Teamwork." In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 13, pp. 11405–11414.

Benke, I., M. T. Knierim, and A. Maedche (2020). "Chatbot-Based Emotion Management for Distributed Teams: A Participatory Design Study." *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW2). DOI: 10.1145/3415189. URL: https://doi.org/10.1145/3415189.

Berendt, B., O. Günther, and S. Spiekermann (2005). "Privacy in E-Commerce: Stated Preferences vs. Actual Behavior." *Communications of the ACM* 48 (4), 101–106. ISSN: 0001-0782. DOI: 10.1145/1053291.1053295. URL: https://doi.org/10.1145/1053291.1053295.

Bittner, E., G. Küstermann, and C. Tratzky (2019a). "The Facilitator is a Bot: towards a Conversational Agent for Facilitating Idea Elaboration on Idea Platforms." In: *Proceedings of the 27th European Conference on Information Systems (ECIS)*.

Bittner, E., M. Mirbabaie, and S. Morana (2021). "Digital Facilitation Assistance for Collaborative, Creative Design Processes." In: *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Bittner, E., S. Oeste-Reiss, P. Ebel, and M. Söllner (2019b). "Mensch-Maschine-Kollaboration: Grundlagen, Gestaltungsherausforderungen und Potenziale für verschiedene Anwendungsdomänen." *HMD Praxis der Wirtschaftsinformatik* 56 (1), 34–49. ISSN: 2198-2775. DOI: 10.1365/s40702-018-00487-1.

Bittner, E. and O. Shoury (2019). "Designing Automated Facilitation for Design Thinking: A Chatbot for Supporting Teams in the Empathy Map Method." In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Brown, T. and B. Katz (2011). "Change by design." *Journal of product innovation management* 28 (3), 381–383.

Carlgren, L., M. Elmqvist, and I. Rauth (2014). "Exploring the use of design thinking in large organizations: Towards a research agenda." *Swedish design research journal* 11, 55–63.

Cautela, C., M. Mortati, C. Dell'Era, and L. Gastaldi (2019). "The impact of Artificial Intelligence on Design Thinking practice: Insights from the Ecosystem of Startups." *Strategic Design Research Journal*.

Chang, C.-W. and J. Heo (2014). "Visiting theories that predict college students' self-disclosure on Facebook." *Computers in Human Behavior* 30, 79–86. ISSN: 0747-5632. DOI: https://doi.org/

10.1016/j.chb.2013.07.059. URL: https://www.sciencedirect.com/science/article/pii/S0747563213002987.

Clawson, V. K., R. P. Bostrom, and R. Anson (1993). "The Role of the Facilitator in Computer-Supported Meetings." *Small Group Research* 24 (4), 547–565. DOI: 10.1177/1046496493244007. URL: https://doi.org/10.1177/1046496493244007.

Cranor, L. F. (2008). "A framework for reasoning about the human in the loop." *Advanced Computing Systems Professional and Technical Association*.

Cranshaw, J., E. Elwany, T. Newman, R. Kocielnik, B. Yu, S. Soni, J. Teevan, and A. Monroy-Hernández (2017). "Calendar.Help: Designing a Workflow-Based Scheduling Agent with Humans in the Loop." In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Denver, Colorado, USA: Association for Computing Machinery, 2382–2393. ISBN: 9781450346559. DOI: 10.1145/3025453.3025780.

Debowski, N., D. Siemon, and E. Bittner (2021). "Problem Areas in Creativity Workshops and Resulting Design Principles for a Virtual Collaborator." In: *PACIS 2021 Proceedings. 108.*

Dellermann, D., A. Calma, N. Lipusch, T. Weber, S. Weigel, and P. Ebel (2019a). "The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems." In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Dellermann, D., P. Ebel, M. Söllner, and J. M. Leimeister (2019b). "Hybrid Intelligence." *Business & Information Systems Engineering* 61 (5), 637–643. ISSN: 1867-0202. DOI: 10.1007/s12599-019-00595-2. URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1543&context=bise.

Elshan, E., D. Siemon, T. de Vreede, G.-J. de Vreede, S. Oeste-Reiß, and P. Ebel (2022). "Requirements for AI-based Teammates: A Qualitative Inquiry in the Context of Creative Workshops." In: *Proceedings of the 55th Hawaii International Conference on System Sciences*.

Hirschprung, R., E. Toch, F. Bolton, and O. Maimon (2016). "A Methodology for Estimating the Value of Privacy in Information Disclosure Systems." *Computers in Human Behavior* 61 (C), 443–453. ISSN: 0747-5632. DOI: 10.1016/j.chb.2016.03.033. URL: https://doi.org/10.1016/j.chb.2016.03.033.

Hjalmarsson, A., J. Recker, M. Rosemann, and M. Lind (2015). "Understanding the Behavior of Workshop Facilitators in Systems Analysis and Design Projects: Developing Theory from Process Modeling Projects." *Communications of the Association for Information Systems* 36.

Holzinger, A. (2016). "Interactive Machine Learning (iML)." *Informatik-Spektrum* 39 (1), 64–68. ISSN: 1432-122X. DOI: 10.1007/s00287-015-0941-6.

Karren, R. J. and M. W. Barringer (2002). "A Review and Analysis of the Policy-Capturing Methodology in Organizational Research: Guidelines for Research and Practice." *Organizational Research Methods* 5 (4), 337–361. DOI: 10.1177/109442802237115. URL: https://doi.org/10.1177/109442802237115.

Knijnenburg, B. P. and A. Kobsa (2013a). "Helping Users with Information Disclosure Decisions: Potential for Adaptation." In: *Proceedings of the 2013 International Conference on Intelligent User Interfaces*. IUI '13. Santa Monica, California, USA: Association for Computing Machinery, 407–416. ISBN: 9781450319652. DOI: 10.1145/2449396.2449448.

Knijnenburg, B. P. and A. Kobsa (2013b). "Making decisions about privacy: information disclosure in context-aware recommender systems." *ACM Transactions on Interactive Intelligent Systems (TiiS)* 3 (3), 1–23.

Kroll, T. and S. Stieglitz (2021). "Digital nudging and privacy: improving decisions about self-disclosure in social networks." *Behaviour & Information Technology* 40 (1), 1–19. DOI: 10.1080/0144929X.2019.1584644. URL: https://doi.org/10.1080/0144929X.2019.1584644.

Laufer, R. S. and M. Wolfe (1977). "Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory." *Journal of Social Issues* 33 (3), 22–42. DOI: https://doi.org/10.1111/

j.1540-4560.1977.tb01880.x. URL: https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4560.1977.tb01880.x.

Leimeister, J. (2014). "Unterstützung der Zusammenarbeit." In: *Collaboration Engineering*. Berlin, Heidelberg: Springer Gabler. DOI: 10.1007/978-3-642-20891-1_3. URL: https://doi.org/10.1007/978-3-642-20891-1_3.

Lembcke, T.-B., S. Diederich, and A. B. Brendel (2020). "Supporting Design Thinking Through Creative and Inclusive Education Facilitation: The Case of Anthropomorphic Conversational Agents for Persona Building." In: *Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference*.

Li, H., R. Sarathy, and H. Xu (2010). "Understanding Situational Online Information Disclosure as a Privacy Calculus." *Journal of Computer Information Systems* 51 (1), 62–71. DOI: 10.1080/08874417.2010.11645450. URL: https://www.tandfonline.com/doi/abs/10.1080/08874417.2010.11645450.

Li, Y. (2012). "Theories in online information privacy research: A critical review and an integrated framework." *Decision Support Systems* 54 (1), 471–481. ISSN: 0167-9236. DOI: https://doi.org/10.1016/j.dss.2012.06.010. URL: https://www.sciencedirect.com/science/article/pii/S0167923612001935.

Liao, M. and S. S. Sundar (2021). "How Should AI Systems Talk to Users When Collecting Their Personal Information? Effects of Role Framing and Self-Referencing on Human-AI Interaction." In: New York, NY, USA: Association for Computing Machinery. ISBN: 9781450380966. URL: https://doi.org/10.1145/3411764.3445415.

Liedtka, J. M. (2015). "Perspective: Linking Design Thinking with Innovation Outcomes through Cognitive Bias Reduction." *Journal of Product Innovation Management* 32, 925–938.

Manheim, K. and L. Kaplan (2018). "Artificial Intelligence: Risks to Privacy and Democracy." *Yale Journal of Law and Technology* 21, 106.

Manikonda, L., A. Deotale, and S. Kambhampati (2018). "What's up with Privacy? User Preferences and Privacy Concerns in Intelligent Personal Assistants." In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. AIES '18. New Orleans, LA, USA: Association for Computing Machinery, 229–235. ISBN: 9781450360128. DOI: 10.1145/3278721.3278773. URL: https://doi.org/10.1145/3278721.3278773.

Masaviru, M. (2016). "Self-Disclosure: Theories and Model Review." *Journal of Culture, Society and Development* 18, 43–47.

McCarthy, J. (2007). "What is Artificial Intelligence?" URL: http://www-formal.stanford.edu/jmc/whatisai/.

Meinel, C. and L. Leifer (2012). "Design Thinking Research." In: *Design Thinking Research: Measuring Performance in Context*. Ed. by H. Plattner, C. Meinel, and L. Leifer. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–10. ISBN: 978-3-642-31991-4. DOI: 10.1007/978-3-642-31991-4_1. URL: https://doi.org/10.1007/978-3-642-31991-4_1.

Metzger, M. J. (2004). "Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce." *Journal of Computer-Mediated Communication* 9 (4). DOI: https://doi.org/10.1111/j.1083-6101.2004.tb00292.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1083-6101.2004.tb00292.x.

Milne, G. R. and M. E. Gordon (1993). "Direct Mail Privacy-Efficiency Trade-offs within an Implied Social Contract Framework." *Journal of Public Policy & Marketing* 12 (2), 206–215. DOI: 10.1177/074391569101200206. URL: https://doi.org/10.1177/074391569101200206.

Patil, S. and A. Kobsa (2005). "Uncovering Privacy Attitudes and Practices in Instant Messaging." In: *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. GROUP '05. Sanibel Island, Florida, USA: Association for Computing Machinery, pp. 109–112. ISBN: 1595932232. DOI: 10.1145/1099203.1099220. URL: https://doi.org/10.1145/1099203.1099220.

Patil, S. and J. Lai (2005). "Who Gets to Know What When: Configuring Privacy Permissions in an Awareness Application." In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: Association for Computing Machinery, pp. 101–110. ISBN: 1581139985. URL: https://doi.org/10.1145/1054972.1054987.

Phillips, P. J., C. A. Hahn, P. C. Fontana, D. A. Broniatowski, and M. A. Przybocki (2020). "Four principles of explainable artificial intelligence." Technical Report. https://doi.org/10.6028/NIST.IR.8312-draft.

Pitt, J. (2012). "Design contractualism for pervasive/affective computing." *IEEE Technology and Society Magazine* 31 (4), 22–29.

Przybilla, L., L. Baar, M. Wiesche, and H. Krcmar (2019). "Machines as Teammates in Creative Teams: Digital Facilitation of the Dual Pathways to Creativity." In: New York, NY, USA: Association for Computing Machinery. ISBN: 9781450360883. URL: https://doi.org/10.1145/3322385.3322402.

Rayo, L. and I. Segal (2010). "Optimal Information Disclosure." *Journal of Political Economy* 118 (5), 949–987. DOI: 10.1086/657922. URL: https://doi.org/10.1086/657922.

Russell, S. and P. Norvig (1995). *Artificial intelligence: a modern approach*. Prentice Hall.

Salim, S. S. and N. M. Yusoff (2008). "Ethics and Information Privacy in Affective Computing." In: *the Proceedings of the Regional Development International Conference and Exhibition (REDICE08)*.

Schallmo, D. R. A. (2017). "Theoretische Grundlagen." In: *Design Thinking erfolgreich anwenden: So entwickeln Sie in 7 Phasen kundenorientierte Produkte und Dienstleistungen*. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 11–28. ISBN: 978-3-658-12523-3. DOI: 10.1007/978-3-658-12523-3_2. URL: https://doi.org/10.1007/978-3-658-12523-3_2.

Seeber, I. (2019). "How do facilitation interventions foster learning? The role of evaluation and coordination as causal mediators in idea convergence." *Computers in Human Behavior* 94, 176–189. ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2018.11.033. URL: https://www.sciencedirect.com/science/article/pii/S0747563218305685.

Seeber, I., E. Bittner, R. O. Briggs, T. de Vreede, G.-J. de Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe, and M. Söllner (2020). "Machines as teammates: A research agenda on AI in team collaboration." *Information Management* 57 (2), 103174. ISSN: 0378-7206. DOI: https://doi.org/10.1016/j.im.2019.103174. URL: https://www.sciencedirect.com/science/article/pii/S0378720619303337.

Siemon, D. and T. Strohmann (2021). "Human-AI collaboration: introducing the virtual collaborator." In: *Collaborative Convergence and Virtual Teamwork for Organizational Transformation*. IGI Global, pp. 105–119.

Starostka, J., M. R. Evald, A. H. Clarke, and P. R. Hansen (2021). "Taxonomy of design thinking facilitation." *Creativity and Innovation Management* 30 (4), 836–844. DOI: https://doi.org/10.1111/caim.12451. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/caim.12451.

Strohmann, T., S. Fischer, D. Siemon, F. Brachten, C. Lattemann, S. Robra-Bissantz, and S. Stieglitz (2018). "Virtual Moderation Assistance: Creating Design Guidelines for Virtual Assistants Supporting Creative Workshops." In: *PACIS 2018 Proceedings. 80.*

Strohmann, T., D. Siemon, and S. Robra-Bissantz (2017). "brAInstorm: Intelligent Assistance in Group Idea Generation." In: *Designing the Digital Transformation*. Ed. by A. Maedche, J. vom Brocke, and A. Hevner. Cham: Springer International Publishing, pp. 457–461. ISBN: 978-3-319-59144-5.

Stutzman, F., R. Capra, and J. Thompson (2011). "Factors mediating disclosure in social network sites." *Computers in Human Behavior* 27 (1). Current Research Topics in Cognitive Load Theory, 590–598. ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2010.10.017. URL: https://www.sciencedirect.com/science/article/pii/S0747563210003158.

Thibaut, J. and H. Kelley (1959). *The social psychology of groups*. John Wiley.

Ulrich, F. (2018). "Exploring Divergent and Convergent Production in Idea Evaluation: Implications for Designing Group Creativity Support Systems." *Communications of the Association for Information Systems* 43. DOI: 10.17705/1CAIS.04306. URL: https://doi.org/10.17705/1CAIS.04306.

Voigt, M. (2014). "Improving design of systems supporting creativity-intensive processes–A cross-industry focus group evaluation." *Communications of the Association for Information Systems* 34 (1), 86.

Wei, K., J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. V. Poor (2020). "Federated Learning With Differential Privacy: Algorithms and Performance Analysis." *IEEE Transactions on Information Forensics and Security* 15, 3454–3469. DOI: 10.1109/TIFS.2020.2988575.

Weinmann, M., C. Schneider, and J. Vom Brocke (2016). "Digital nudging." *Business & Information Systems Engineering* 58 (6), 433–436.

Westlund, J. K., J. J. Lee, L. Plummer, F. Faridi, J. Gray, M. Berlin, H. Quintus-Bosz, R. Hartmann, M. Hess, S. Dyer, et al. (2016). "Tega: a social robot." In: *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 561–561.

Xiao, Z., M. X. Zhou, and W.-T. Fu (2019). "Who should be my teammates: Using a conversational agent to understand individuals and help teaming." In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 437–447.

Xu, H., T. Dinev, H. Smith, and P. Hart (2008). "Examining the Formation of Individual's Privacy Concerns: Toward an Integrative View." In: *ICIS 2008 Proceedings. 6*.

Xu, H., H.-H. Teo, B. C. Y. Tan, and R. Agarwal (2009). "The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services." *Journal of Management Information Systems* 26 (3), 135–174. DOI: 10.2753/MIS0742-1222260305. URL: https://doi.org/10.2753/MIS0742-1222260305.

Youn, S. (2009). "Determinants of Online Privacy Concern and Its Influence on Privacy Protection Behaviors Among Young Adolescents." *Journal of Consumer Affairs* 43 (3), 389–418. DOI: https://doi.org/10.1111/j.1745-6606.2009.01146.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-6606.2009.01146.x.

Zoom Video Communications, I. *Attendee attention tracking*. https://support.zoom.us/hc/en-us/articles/115000538083-Attendee-attention-tracking. Accessed: 07-27-2021.