Let the computer evaluate your idea: evaluation apprehension in human-computer collaboration

Siemon Dominik

DOI: 10.1080/0144929X.2021.2023638

Let the Computer Evaluate Your Idea – Evaluation Apprehension in Human-Computer Collaboration

Individuals tend to hold back their ideas because they feel concerned about being evaluated. This leads to the untapped creative potential for organizations that depend on the creative abilities and ideas of their employees, as idea evaluation is essential for further developing and assessing creative ideas that inhibit the potential to turn into innovative products or services. In our research, we propose the use of AI-based computer systems for idea evaluation to address evaluation apprehension. With the help of an experiment (n=228), we test whether individuals feel concerned about evaluation when a computer evaluates their idea. Our results show that people do not feel evaluation apprehension when they present their idea to an AI-based system, but in contrast, feel concerned when they present their idea to a human. These findings contribute to the theory of evaluation apprehension but also to theories of human-computer collaboration and hold potential for companies to increase their creative outcome.

Introduction

Ambitious customer needs, newly emerging technologies, and an ever more connected society are putting companies in a constant position to produce and establish innovations, for which teamwork and effective collaboration has proven to be beneficial (Cui & Wu, 2016; Pisano, 2015; Yan et al., 2018). The ability to come up with, appraise and further refine new and innovative ideas, involves a range of creative activities that, when deployed in a team, can unfold its full potential (Amabile & Pratt, 2016; Somech & Drach-Zahavy, 2013). In this context, the use of creativity techniques and effective collaboration are essential for the development of products and services, which, with the help of information and communication technology, can take place anywhere and at any time (Hoever et al., 2012; X. Wang et al., 2015).

One established and well-investigated approach to support creativity is brainstorming, which on the one hand has shown to be beneficial when it is important to come up with many ideas. On the other hand, brainstorming leads to negative effects, such as production blocking and evaluation apprehension, that inhibit the overall creativity of the participants (Barki & Pinsonneault, 2001; Camacho & Paulus, 1995; Dennis & Williams, 2003). Evaluation apprehension is a major inhibitor when it comes to collaborative creativity (Zhou et al., 2019), as individuals "may not want to put forth wild ideas if they are afraid of losing credibility, having their idea rejected, or being humiliated" (Wilson, 2013, p. 129). Studies have shown that social aspects, such as hierarchy, are factors that influence individual evaluation apprehension and subsequently lead to untapped creative potential and possible missed innovations (Zhou et al., 2019). Even though, organizations possess vast creative potential, fostering and encouraging creative contributions to overall enhance organizational effectiveness is often neglected (DiLiello & Houghton, 2008). Against this background, computer systems are increasingly being used that are specifically designed to promote creativity, ideation generation, as well as to allow ideas to be evaluated by other people (K. Wang & Nickerson, 2017). While these systems primarily serve to enable the user to enter and process ideas, technological progress, especially in the field of artificial intelligence (AI) is creating further opportunities to involve computer systems in a more intelligent and active way for example by generating or evaluating ideas (Siemon et al., 2022). Because of the increase in computing power, novel algorithms and faster, more available data have contributed considerably to AI maturing (Russell & Norvig, 2020). With this increase of computational intelligence, natural language processing becomes more capable, enabling an efficient and powerful human-computer interaction and collaboration (Aleksander, 2017; de Vreede & Briggs, 2019; Diederich et al., 2022; Porra et al., 2019; Seeber et al., 2020). This has led to a recent increase in applications such as virtual assistants like IBM's Watson, Apple's Siri or Amazon's Alexa within the public sector and systems like enterprise bots that are applied within organizations (Diederich et al., 2022; Lee et al., 2019; Maedche et al., 2016; Stieglitz et al., 2018). AI is a versatile and often overused term that applies to several concepts, such as machine learning, neural networks, robotics and more (Russell & Norvig, 2020). In our research, and especially in the context of human-computer interaction, we define and use the term AI-based computer systems, which describe intelligent systems such as conversational agents or virtual assistants that interact with humans in a natural and intelligent way.

However, these AI-based computer systems mainly focus on assisting activities by passively serving users within their work (Diederich et al., 2022; Gnewuch et al., 2017; Rzepka & Berger, 2018). By increasing the autonomy and independence of AI-based computer systems, their role within value creation changes (Siemon et al., 2022). For example, implementing such systems in collaboration scenarios, where the system has an active role in the collaboration process, results in computers becoming teammates (Gerber, 2009; Krämer et al., 2012; Nass et al., 1996; Seeber et al., 2019, 2020; Siemon et al., 2018). Subsequently, "computers have now taken on roles that go beyond being mere tools" (Nass et al., 1996, p. 669). Such hybrid intelligence, where humans and computers complement each other, can be used for organizational decision making (Dellermann et al., 2019; Jarrahi, 2018) and creativity-intensive processes, where creativity is needed to come up with novel ideas (Aleksander, 2017; Anderson et al., 2018; Colton & Wiggins, 2012).

Computers performing creative tasks (referred to as computational or artificial creativity) is considered as the final frontier of AI and describes a multidisciplinary endeavor of creating AI-based systems that are capable of human-level creativity (Besold et al., 2015;

Colton & Wiggins, 2012). Already in the proposal that led to the famous Dartmouth conference, the term creativity was mentioned among other things as a key goal of AI (Boden, 2009; McCarthy et al., 2006). Regardless of whether people work together with computers or computers generate value on their own, new approaches on how AI can be involved in creative processes ensure that not only the capabilities but also the role of computers must be reconsidered and above all researched (Besold et al., 2015; Siemon et al., 2022). In this paper, we propose the use of AI-based computer systems that actively participate in creative processes by evaluating a user's idea and thus providing feedback, as an approach to address evaluation apprehension.

When an AI-based computer system interacts and even collaborates with humans (Loebbecke & Picot, 2015; McTear et al., 2016; Spinella, 2018), research is challenged to address and possibly refine existing theories on collaboration mechanisms and social factors in creative teamwork. Current research already provides evidence on how humans interact with AI-based computer systems (Nass et al., 1996) and how specific social mechanisms can be applied to the same extent as to human-human interaction (Nass & Moon, 2000). If an AI-based computer system visually appears human-like or shows human-like behavior, individuals tend to mindlessly apply social rules to computers and project onto them characteristics that are normally associated with humans (Nass & Moon, 2000). In creative tasks, such systems can be used to evaluate ideas in the early stages, where support is essential to further refine and develop an idea (Mumford et al., 2012). Especially in the early stages of idea generation, evaluation apprehension is most present, as partially developed concepts can be misunderstood or criticized due to their incompleteness (Diehl & Stroebe, 1987). However, only limited research exists (Lubart, 2005; Siemon et al., 2015), whether humans experience evaluation apprehension when working with AI-based computer systems on creativity-intensive tasks. Therefore, we aim to answer the research question:

RQ: To what extent do humans experience evaluation apprehension when AI-based systems evaluate their idea?

We address this research question with an online experiment, where we simulated an AI-based idea evaluation system called "Alan" that assesses and comments on a user-generated idea compared to a simulated human that evaluates an idea to the same extent. Prior to the idea generation, the participants fill out a survey to indicate their general fear of being evaluated, after which they are asked to write an idea that will further be evaluated. At this stage, the participants are introduced to the AI-based system or the human that is going to evaluate their idea. At this point, the participants are asked about their fear of being evaluated by that specific AI-based system or simulated human in order to find out the actual evaluation apprehension. The results indicate that participants that are introduced to the AI-based computer system feel significantly less evaluation apprehension than the participants introduced to the human idea evaluator.

Theoretical Background

Creativity and Evaluation Apprehension

Creativity is an intangible term that can be associated with an outcome of a process, a person or a group, a process itself, or a way of thinking (Hennessey & Amabile, 2010; Runco & Jaeger, 2012). A well-accepted definition by Amabile (1983) says, that a creative outcome is novel and appropriate, useful, correct or valuable to a specific task. Also, creativity is an activity that is heuristic in nature and involves a process (Mumford et al., 2012). Creativity is divided into four different components; the creative process referred to as ideation, the idea as the output of the creative process, the individual carrying out the process and the environment around the process and the person (Basadur et al., 2000; Rhodes, 1961). In order to generate an idea, the creative process involves a variety of activities, such as different ways of thinking, the forming of associations and the combination of existing knowledge with the ability to perform mental leaps into different and novel dimensions of knowledge (Mumford et al., 2012). Researchers argue that this process consists of an interplay of divergent thinking, the process used to generate creative ideas by exploring many possible solutions, and convergent thinking, the process of fleshing out many ideas or one idea and getting a well-crafted solution (Mumford et al., 2012; Runco, 2003). Analogously, an idea is thus generated and evaluated again and again, i.e. expanded and (re)invented, as well as evaluated and improved (Hennessey & Amabile, 2010; Mumford et al., 2012). Idea generation pursues thereby the goal to form new associations, create novel thoughts and convert them into original ideas (Baer, 2014). While during idea evaluation these thoughts need to be evaluated according to their novelty and originality, as well as their feasibility, acceptability and effectiveness (Dean et al., 2006). During the creative process, the current state of the idea is constantly evaluated, refined and further developed, leading to a continuous improvement of the idea (Santanen et al., 2004). Thereby, the process of idea generation and idea evaluation do not have to take place consecutively, but can also happen simultaneously. Idea generation and evaluation are therefore two interlocking processes that often run synchronously (Runco, 2003).

When generating ideas individually, these activities are restricted to the single individual that creates the idea and the environment that the individual is in (K. Wang & Nickerson, 2017). In collaborative creativity, the environment is extended with other individuals that influence the collective performance with comments, further information, their ideas and other input, which are responsible for further enhancing existing ideas (Barczak et al., 2010; Fink et al., 2012; Mamykina et al., 2002). This support is not solely due to different abilities, knowledge, and skills by each individual, but because of so-called synergetic effects that go beyond the sole sum of each individual's performance (Paulus et al., 2012; Tasa et al., 2007). Those synergetic effects influence team performance and can support the creative process by fostering a mutual motivation and group awareness, but they also impair it due to cognitive overload, production blocking, social loafing or evaluation apprehension (Diehl & Stroebe, 1991; Karau & Williams, 1993; Zhou et al., 2019). Nevertheless, it is important to transform the initially generated thoughts into a mature and well-developed idea, even if negative synergistic effects exist due to the fear of criticism. Further refinement of the idea during or after ideation is of great importance to reach the full potential of an initial thought (Hung et al., 2017; Licuanan et al., 2007; Runco, 2003; Sawyer & DeZutter, 2009).

For organizations using systematic innovation and idea management with internal employees and external users, the development and improvement of ideas is an essential aspect to consider before generated ideas are pursued further (Yan et al., 2018). Only mature ideas are presented to relevant decision-makers which can only then be pursued with financial resources. Therefore, it is important to evaluate ideas to correct the necessary weaknesses and recognize their potential (Licuanan et al., 2007). Besides the importance of idea evaluation, whether consecutively after an initial thought was generated or during ideation, the process often involves evaluation apprehension which causes individuals to hold back their ideas, and so subsequently, there is untapped creative potential (Zhou et al., 2019). Not only do ideas need to be evaluated within companies in order to improve them, but also in the area of crowdsourcing, where external stakeholders are used to contribute ideas or evaluate ideas (Zhao & Zhu, 2014). The crowd can thus be used to apply its extensive knowledge to ideas by evaluating individual ideas. In this context, too, evaluation apprehension prevails, since users who contribute their ideas are either afraid to put their ideas up for public evaluation or to have them evaluated by a committee (Kim et al., 2010).

The theory of evaluation apprehension states that participants who are working in groups or teams are not presenting their more original ideas, because of the fear of negative evaluations from other members of the group (Diehl and Stroebe, 1987). The theory was first introduced in 1972, by Nickolas B. Cottrell, who argued that social aspects in the form of approval or disapproval by other humans are based on their assessment. The mere presence of others is already sufficient to produce arousal that triggers evaluation apprehension and distracts an individual within a creative process (Cottrell et al., 1968; Karau & Williams, 2001). Due to aspects such as losing credibility, having an idea rejected, or being humiliated, individuals tend to hold back wild, unfinished or unpopular ideas that however can be refined and further developed into valuable ideas (Zhou et al., 2019). When an individual is surrounded by experts or superiors, evaluation apprehension is especially severe because of hierarchical structures (Zhou et al., 2019). Evaluation apprehension is thus a variable that can be manipulated by creating different situations that can lead to fear of evaluation, such as surrounding an individual by experts, close friends or by the perception of being publicly judged (Kim et al., 2010). However, evaluation apprehension is also influenced by individual character traits such as self-esteem, social desirability or anxiety and is therefore pronounced differently in people (Thomas et al., 1979). People with low self-esteem, anxiety or high social desirability tend to have higher evaluation apprehension. Since self-esteem and anxiety are personal characteristics that are constantly changing, fear of negative evaluation is also a changing personality trait that however can be manipulated very effectively. In summary, evaluation apprehension is an individual characteristic that manifests differently in each and every person and correlates with other psychological constructs such as self-esteem, anxiety and social desirability. However, evaluation apprehension can also be altered leading to different apprehensive feelings of being evaluated in individuals (Huang et al., 2015; Simons & Turner, 1976).

Research on evaluation apprehension is often conducted in the context of brainstorming and electronic brainstorming, where criticism shouldn't be expressed during ideation (Dennis & Williams, 2003). With the rule of not expressing criticism and not evaluating ideas generated by others, brainstorming makes a fundamental demand to address evaluation apprehension. Nevertheless, especially in in-person brainstorming sessions, criticism can be expressed indirectly through facial expressions and body language (Dennis & Williams, 2003; Diehl & Stroebe, 1987). Ideas should be built upon and further refined, which means that ideas in a team must be presented to everyone. Since individuals already tend to hold back their ideas due to the presence of others, researchers and practitioners suggest mechanisms such as anonymity to manipulate evaluation apprehension, which can be particularly well implemented in digital scenarios (Connolly et al., 1990; Nunamaker et al., 1991; Shepherd et al., 1995). When contributing anonymously, individuals are encouraged to be more open, honest and free in their participation and are willing to submit even unconventional and unpopular ideas for discussion and rating (Connolly et al., 1990; Jessup et al., 1990). However, when individuals are anonymized, negative effects such as social loafing or a harsh and disrespectful evaluation can occur (Chen et al., 2014; Connolly et al., 1990; Rao, 1992). Individuals tend to social loaf when they are not engaged in contributing, and general reciprocal behavior is undermined because of anonymity (Nunamaker et al., 1987; Shepherd et al., 1995). In addition, due to anonymity, comments tend to be more harsh and disrespectful, leading to an overall hostile evaluation (Wright, 2013; Zhou et al., 2019). Subsequently, there is no clear recommendation and guideline for the use of anonymity, as it may address issues such as evaluation apprehension, but also leads to various performance losses. The effect of anonymity in creative processes is not consistent, as some studies have found no effects of anonymity (Jessup et al., 1990; Sosik et al., 1998) while others depict that anonymity encourages individuals to engage in social loafing and free-riding, leading to reduced team performance (Shepherd et al., 1995). The use of anonymity is therefore controversial, and other mechanisms and functionalities that support collaborative creativity are used primarily in digital creativity support.

Artificial Intelligence in Creativity

With the help of information technology, a variety of methods for supporting creativity have already been digitized (Gabriel et al., 2016; K. Wang & Nickerson, 2017). These digitized techniques aim to further support the creative process, by using different mechanisms and functions that information technology offers (Massetti, 1996). Computer programs or systems are capable of aiding users to complete tasks in various fields, ranging from engineering tasks and medical services to decision support and knowledge management support (Kraemer & King, 1988; Power & Sharda, 2009). Systems that have been designed for tasks in which creativity is necessary, including science studies, literary work, musical work, musical depiction, design, mathematics, and art (Gabriel et al., 2016) are called creativity support systems (CSS).

Lubart (2005) divided the support provided by a CSS into four categories and figuratively named them: Computers as a nanny, as a pen-pal, as a coach or as a colleague. The nanny supports by monitoring the creative process, setting agendas and deadlines. The pen-pal helps by providing the possibility to receive, compose and distribute an idea; the coach represents a support system that helps with a specific method or with the system itself, e.g. by recommending other methods or steps within idea generation, and the computer as a colleague depicts an AI-based system that actively takes part in a creative process (Lubart, 2005). Lubart et al. (2005, p. 368) argue, that “it is possible to conceive of computers as real partners in the creative process intervening at different points in order to generate, evaluate, or refine ideas and bring them to full-fledged products.” The field of computational creativity, studies the building of such intelligent systems that exhibit creative behavior that is similar to humans (Colton & Wiggins, 2012).

Due to the rapid increase in computing power and the exploration of new methods and algorithms, AI is now capable of forming associations and making mental leaps. What initially began as theoretical derivations and models, it is now possible (Besold et al., 2015; Boden, 1998; Colton & Wiggins, 2012). Abilities that could initially only be attributed to living beings and especially humans can now be performed by AI (Besold et al., 2015). Although many researchers see creativity as something intangible and even mystical (Boden, 2004), others see creativity as something systematic and explainable, which means that "computers can- and do- exhibit the same kinds of behaviors that creative humans do" (Colton et al., 2009, p. 12). As one of the leading researchers in computational creativity, Margaret Boden argues that research on computational creativity in-fact, helped towards a better understanding of creativity and that combinatorial and transformational exploration can be performed by computers (Boden, 2009). With said creative systems, individuals could use programs for creative tasks where the computer extends the human’s cognition and acts as a creative collaborator e.g. by evaluating, commenting and providing further feedback to an idea.

For instance, Maher and Fisher (2012) developed an AI-based system that evaluates the novelty, the degree of surprise, the unexpectedness and the value of an idea. With the help of AI-based clustering methods, the distances of the key figures to other products are evaluated for the criterion novelty. The degree of unexpectedness is evaluated by comparing the development patterns, whereas, for the value of the idea, an adaptive function with a genetic algorithm is used (Maher & Fisher, 2012). A study by Varshney et al. (2019) deals with the automatic evaluation of a pool of ideas in the domain of culinary recipes. The evaluation of the novelty of recipes from a generated pool is done using the Bayesian Surprise method. Other approaches for AI-based idea evaluation that exist rely on methods like latent semantic analysis, latent dirichlet allocation and term frequency-inverse document frequency. Such AI-based evaluation methods can be compared to human expert evaluation (K. Wang et al., 2019).

If such AI-based systems now take on a special role, task or activity like idea evaluation, they become equal partners and a collaboration scenario arises (Seeber et al., 2019; Siemon et al., 2018). For example, Hwang & Won (2021) found that “that participants consistently contributed more ideas and ideas of higher quality when they perceived their teamworking partner to be a bot” (p. 1). This leads to an active involvement in a collaborative creativity performance where theories of human-AI collaboration must be considered (Aleksander, 2017; Daugherty et al., 2018; Nass et al., 1996; Seeber et al., 2019; Siemon et al., 2018; Strohmann et al., 2017).

Computers are Social Actors

When generally interacting with media, individuals tend to treat computers and other media as if they were humans (Reeves & Nass, 1996). This leads to the hypothesis that established communication and interaction theories on how humans behave and respond in human-human scenarios can be applied in the same extent to a human-AI scenario because “individuals’ interactions with computers, television, and new media are fundamentally social and natural, just like interactions in real life” (Reeves & Nass, 1996, p. 5).

In a study from 1996, subjects were interdependently working with a computer in a team to create team affiliation with group dynamics like in human-human interaction. The results showed effects of being in a team, whereas the subjects experienced influence from the computer similar to group dynamics as in human-human teams (Nass et al., 1996). By simply telling the subjects that they are working with a computer, they affiliate with computers in a team relationship and "display the same sorts of attitudes and behaviors as when working in teams with other humans" (Nass et al., 1996, p. 675). Many studies provide empirical evidence that computers can induce social reactions similar to those in human-human interaction scenarios (Rickenberg & Reeves, 2000; Sproull et al., 1996). Besides increased cooperative behavior (Parise et al., 1999), social inhibition can occur in human-AI interaction, leading to an impaired collaboration. Experimental studies showed that individuals monitored by animated agents during individual tasks, feel anxiety leading to decreased task performance (Rickenberg & Reeves, 2000). Observations also showed that the presence of a robot during the performance of a task causes apprehension and generally induces social facilitation effects (Woods et al., 2005). The assumption seems obvious that humans mindlessly apply social responses towards computers in the same way that they do towards humans (Nass & Moon, 2000). Nass and Moon (2000) provide insights from a series of experimental studies that led to the origin of the computers are social actors paradigm with observations of anthropocentric reactions to computers. One aspect that has been studied in more detail is reciprocity, a social phenomena which means that individuals are willing to help or contribute when a team member also actively contributes (Kankanhalli et al., 2005). When a computer contributes, it triggers the mindless response of the individual feeling obliged to contribute as well, which is a major aspect of collaborative creativity (Nass & Moon, 2000). Although studies have shown that traditional group dynamics, social responses, and general interaction theories can be applied to human-AI interaction, as there are "more similarities between human-human and human-machine interactions than differences" (Krämer et al., 2012), still, not all phenomena of human-AI interaction have yet been investigated.

In order to develop individual ideas further, they have to be presented, evaluated and extended by others (DiLiello & Houghton, 2008). The aspect that one's own, possibly wild and unfinished ideas are evaluated and improved by others, leads to individuals feeling concerned and thus they tend to hold back their ideas (Zhou et al., 2019). Therefore, when people create an idea and subsequently present it to another person for evaluation, the fear of negative evaluation arises (Huang et al., 2015; Zhou et al., 2019). If this real person is replaced by an AI-based computer system, the question arises to what extent fear of evaluation is present. A fundamental property is that an AI-based system first reveals its identity to the user, in order to reach the perception of being an AI-based system (Luo et al., 2019). Theories of human-computer interaction and collaboration state that humans respond and behave similar to AI-based systems that possess human-like behavior, as to real humans (Isbister & Nass, 2000; Krämer et al., 2012; Nass & Moon, 2000). This however depends on the fact that the AI-based system possesses and shows human-like behavior and that the individuals perceive it accordingly (Qiu & Benbasat, 2009). Furthermore, AI-based systems with human-like characteristics can also evoke feelings of social presence, which emerges when interacting with human-like systems that transmit any kind of social cues (Verhagen et al., 2014). These indicators (i.e. social cues) can vary greatly and differ in verbal aspects such as giving tips or advice, as well as joking or excusing (Feine et al., 2019). In addition, visual cues (e.g. appearance) or invisible cues (e.g. response time) invoke social presence. Possessing a number of these cues, leads to the perception of humanness and social presence, which are both essential aspects of the computers are social actors theory (Qiu & Benbasat, 2009).

Hypothesis Development

It is important to design the AI-based computer system that evaluates the idea in such a way that users do not perceive it as human and also do not feel any social presence. Consequently, if an AI-based system is perceived as human and users feel social presence, it can be hypothesized that they will also feel evaluation apprehension when the system evaluates their idea. However, a system that is clearly perceived as an AI-based system should not elicit this. To measure whether the participant's evaluation apprehension is affected by the idea evaluator (human or AI-based computer system), we need to measure the perceived humanness and social presence of both simulated idea evaluators. Therefore, perceived humanness and social presence are relevant preconditions for our manipulation, which is why we constitute a manipulation check (**MC**).

MC: There is a significant difference in perceived humanness and social presence between the AI-based computer system and the human.

Only if the manipulation is successful, we can continue with our first hypothesis. When the AI-based computer system takes over the role of idea evaluation, it can be assumed that people do not fear being evaluated, when it is perceived as less human and if the degree of social presence is less. Thus, we propose the following hypothesis:

H1: There is a significant difference in evaluation apprehension between individuals presenting their idea to an AI-based computer system and individuals that present their idea to a human.

At this point, it is important to mention that perceived humanness and social presence are preconditions for our hypothesis, but cannot serve as predictors for evaluation apprehension. The degree of perceived humanness and social presence does not influence the degree of evaluation apprehension, as evaluation apprehension is an individual trait (i.e. social anxiety) that develops over time based on experience with previous criticism and evaluation (Carleton et al., 2007; Leary, 1983; Zhou et al., 2019). In other words, a more human idea evaluator does not lead to less evaluation apprehension, but a human idea evaluator leads to evaluation apprehension. Thus, we define social presence and perceived humanness as measures for the manipulation check and not as predictors. Against this backdrop, evaluation apprehension is a characteristic that can be of varying in intensity (Carleton et al., 2007). People with low evaluation apprehension feel no or only limited fear of criticism, whereas people with high evaluation apprehension are generally very much deterred by a potential evaluation and are very likely to hold back their idea (Carleton et al., 2007; Zhou et al., 2019). As people with low evaluation apprehension do not fear being evaluated by other persons, we hypothesize that they also do not fear being evaluated by an AI-based system (H2a). However, of particular interest are individuals that fear being evaluated and therefore have a medium or high level of evaluation apprehension in general. In this case, an AI-based system would be beneficial to overcome evaluation apprehension, foster idea generation and thus reduce potentially missed ideas. Therefore, we hypothesize that individuals with medium (H2b) and high (H2B) evaluation apprehension do not fear being evaluated, when presenting their ideas to an AI-based system but feel being criticized when presenting their idea to a human.

H2a: There is no significant difference in evaluation apprehension between individuals presenting their idea to an AI-based computer system and individuals that present their idea to a human for people with a low value of evaluation apprehension.

H2b: There is a significant difference in evaluation apprehension between individuals presenting their idea to an AI-based computer system and individuals that present their idea to a human for people with a medium value of evaluation apprehension.

H2c: There is a significant difference in evaluation apprehension between individuals presenting their idea to an AI-based computer system and individuals that present their idea to a human for people with a high value of evaluation apprehension.

Furthermore, it is important, whether the use of an AI-based system can address evaluation apprehension for those that fear being evaluated. The objective of this study is, if an AI-based system can be utilized for evaluating ideas by lowering the fear of being criticized for individuals with a medium or high level of evaluation apprehension. Therefore, we hypothesize that individuals in the treatment condition (presenting their idea to an AI-based system) with a medium and high level of evaluation apprehension in general have a significant decrease in evaluation apprehension, when presenting their idea to an AI-based system.

H3a: There is a significant difference between the general (i.e. individual) fear of being evaluated and the experienced evaluation apprehension when presenting their idea to an AI-based system for people with a medium value of evaluation apprehension.

H3b: There is a significant difference between the general (i.e. individual) fear of being evaluated and the experienced evaluation apprehension when presenting their idea to an AI-based system for people with a high value of evaluation apprehension.

H3c: There is no a significant difference between the general (i.e. individual) fear of being evaluated and the experienced evaluation apprehension when presenting their idea to an AI-based system for people with a low value of evaluation apprehension.

To better understand the hypothesis, figure 1 presents the conceptual model of the developed hypotheses.

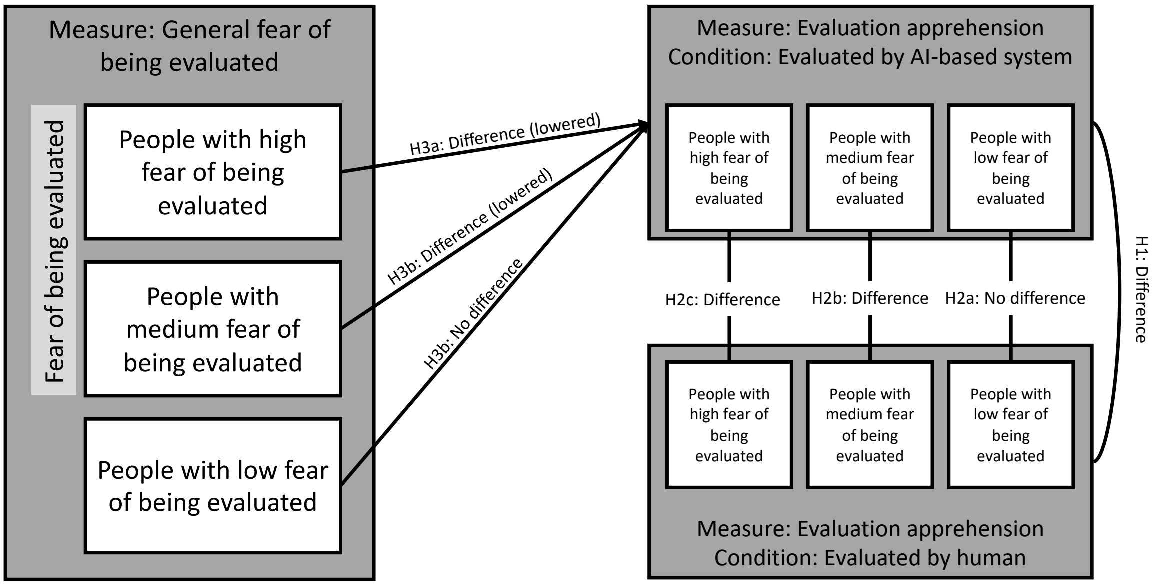


Figure 1. Conceptual model and hypothesis visualization

Methodology

In order to address our research question and test our hypotheses, we conducted an online experiment. We chose an experimental study to maximize internal validity (Karahanna et al., 2018) and to be able to validate if any observed differences in experienced evaluation apprehension, i.e. the dependent variable are caused by our manipulation, i.e. the evaluation by a human or an AI-based system (independent variable) (Brewer & Crano, 2000). We followed the Wizard of Oz approach (Riek, 2012), in which the intelligent system is simulated to suggest the appearance of an AI to the subject. This type of experiment dates to 1789 and is often utilized in studies involving interaction with AI-based systems (Dahlbäck et al., 1993; Riek, 2012). In our experimental design, a human (the subject) assumes to interact with an autonomous (in the sense of AI) system, to assess his or her perception, reactions, and behaviors towards the AI-based system.

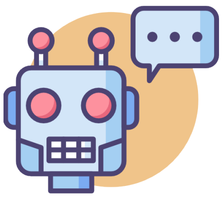
A total of 228 people (male = 115, female = 100, diverse = 1, no answer = 2) took part in our experiment, which was distributed through personal contacts, student mailing lists, social media and online forums. More than half of the participants (56.8%) are between 18 – 29 years old, whereas 19.47% are between 30 – 39, 5.33% between 40 – 49, 3.20% between 50 – 59, 1.33% over 60 and 1.07% under 18. 0.80% did not provide information about their age. The highest educational degree of 40.80% of the participants is a university degree, whereas of 18.13% it is a college degree. 12.53% have a polytechnic degree, 6.93% have an apprenticeship, 1.87% have a doctorate, whereas 6.93% do not have a degree and 14,68% did not answer this question. Participants are mostly students (39.20%) or are working in fields such as social care, commercial services, trade, distribution, hotel and tourism, economics, computer science, logistics, security, administration or others. Participation was not mandatory and was not remunerated. The experiment took about 15 minutes to complete.

Experimental Design and Procedure

Our experiment follows a between-subject design, where participants experience either the experimental (EG) or the control condition (CG) (große Deters et al., 2019). As several characteristics of the participants can introduce variation (Shaughnessy et al., 2000), an automated randomization was performed by a PHP function (*Math.random()*) in the simulation, resulting in random distribution of participants to either the experimental group (EG, n =117) or the control group (CG, n = 111). The procedure of the experiment is as follows.

After a greeting of the participants and a short introduction to the upcoming task, the participants were asked to conduct an initial survey. This survey includes demographics and the BFNE-R (brief fear of negative evaluation scale – revised) inventory, which measures the general fear of being evaluated (Carleton et al., 2007). The participant is then randomly introduced to a partner who is either a simulated AI (EG) or a simulated human (CG). The AI-based system is named Alan and introduces itself as a creative agent and thus reveals its identity as being an AI-based system (Luo et al., 2019). The human partner is named Phillip Stoecker, which is a generic name, produced randomly using a name generator. Alan has an avatar based on an openly available icon, representing an AI-based system, whereas Philipp Stoecker has a picture from a stock picture website. To control extraneous variables and observe a true effect of the manipulation, except for their names, both introduced themselves identically and explained their task and role in the experiment. To simulate a connection establishment with the real human (Philipp Stoecker), a short delay of 5.5 seconds was built in. Table 1 shows how Alan and Philipp Stoecker introduced themselves to the participants.

Table 1. Introduction of Alan and Philipp Stoecker

 Alan	 Philipp Stoecker
"Hello, I'm Alan. I am a creative agent and will support you with comments, feedback and a subsequent evaluation of your idea. I was programmed to understand and evaluate your idea. I will pay special attention to how innovative and feasible your idea is."	"Hello, I'm Philipp Stoecker. I will support you with comments, feedback and a subsequent evaluation of your idea. I will pay special attention to how innovative and feasible your idea is."

After this introduction, and before the participants started to work on their idea, they were again asked to carry out a short survey. This survey contains the BFNE-R inventory, which is tailored to the corresponding process of idea generation and idea evaluation necessary in the subsequent step. After this second survey, the participants were asked to write down an idea for a task. The idea must have a minimum length of 400 characters, which ensures that the participant deals with the problem and develops a sufficient idea. A generally known problem was chosen as the task, which does not require special expertise and requires as little personal involvement as possible. The following task was presented to the participant:

"Mobility means freedom, independence and self-determination but also a lot of problems for people and the environment. Think about a new and innovative idea for a sustainable mobility of the future. This could be a product (e.g. innovative vehicle), a service or a new type of process. Be creative!"

Again, both Alan and Philipp Stoecker used the same words in this step. After the idea generation, the idea is sent and an evaluation of the idea takes place. With Philip Stoecker, an artificial delay of a of total 175 seconds is built-in (i.e. the average time to read an idea of 400 characters in length and formulate a series of comments and evaluations), simulating the activity to read the idea and work out an evaluation. In Alan's case, only a delay of 4 seconds is built-in. Thereupon an evaluation with numerous comments is displayed which is a generic evaluation that addresses different general aspects of an idea and identical for both, Alan and Philipp Stoecker (see Appendix). After that, a final survey was carried out including a manipulation check where participants assessed the perceived humanness (PH) and social presence (SP) of Alan or Philipp Stoecker as well as further open questions (OQ) on AI in idea evaluation. Figure 2 provides an overview of the experimental design.

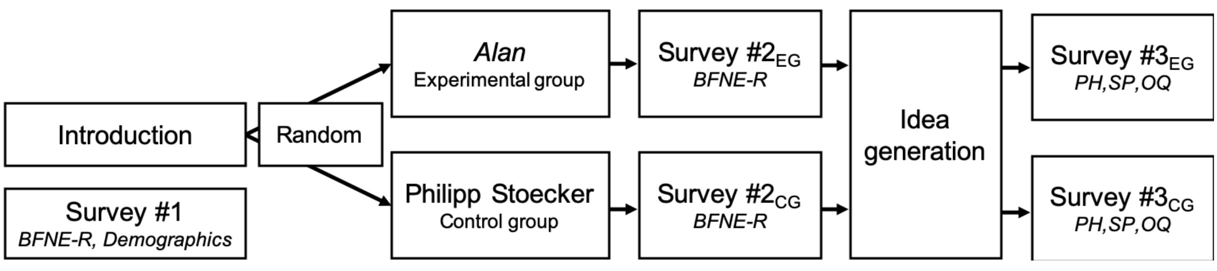


Figure 2. Overview of the experimental design

Measures

In 1969, Watson and Friend developed the fear of negative evaluation scale (FNE), a 30-item self-rated questionnaire to measure evaluation apprehension. This widespread scale measures the feeling of concern about other people's evaluation; the disappointment with these negative evaluations and the expectation that others will leave negative evaluations (Watson & Friend, 1969). The scale relates to the feeling of unease of being evaluated "unfavorably while anticipating or participating in a social situation" (Hofmann & DiBartolo, 2014, p. 455). As its utility is limited by length, Leary (1983) developed the Brief-FNE, a reliable 12-item scale representing a workable alternative to the FNE, that was later revised by Carleton et al. (2007) resulting in the BFNE-R scale. In our study, we used the German version of the BFNE-R (called FNE-K[1]) that was translated and validated with four different studies in 2015 by Reichenberger et al. "The scale demonstrated appropriate test-retest reliability, convergent and discriminant validity, criterion validity" (Reichenberger et al., 2015, p. 169) and was therefore used to assess evaluation apprehension in the experiment. We used the measure two times during the experiment; at the beginning to assess the general fear of negative evaluation[2] and directly before the idea generation to assess the fear of (possible) negative evaluation[3] by Alan or Philipp Stoecker (see Figure 2). The measure was answered on a 5-point Likert scale (1 = low, 5 = high).

---

[1] We will continue to use the English abbreviation, BFNE-R

[2] We use *general BFNE-R* for the general fear of being evaluated of our participants

[3] We use *BFNE-R* for the evaluation apprehension the participants experience prior to their idea generation and idea evaluation

A manipulation check was included within the last survey to ensure how human the participants perceived Alan and Philipp Stoecker and how they rated the social presence of both. We used the perceived humanness measure (6 items) by Holtgraves & Han (2007), which was measured on a 7-point Likert scale (1 = high, 7 = low) and the social presence measure (5 items) (Gefen & Straub, 2004), which was also measured on a 7-point Likert scale (1 = high, 7 = low) as used by Qiu & Benbasat, (2009). Furthermore, open questions regarding the idea evaluation by Alan (only in the experimental condition) and the overall perception of the idea evaluation were asked (both groups). All measures can be found in the Appendix.

Data Analysis and Results

Data was analyzed by means of descriptive analyses, confirmatory factor analysis and a series of Mann-Whitney-U tests to understand if our manipulation led to a significant difference in our dependent variable. All analyses were carried out using the statistical computing software RStudio (Version 1.2.5033). Due to non-normally distributed data (see Table 3) and the use of ordinal scales (Wu & Leung, 2017), we chose the nonparametric Mann-Whitney-U test for our two independent groups. We used confirmatory factor analysis to test our measures, which showed that all items have significant (significance level of $p \leq .001$) positive factor loadings ranging from .611 to .930. Furthermore, we calculated composite reliability (CR) and average variance extracted (AVE), both indicating reliable factors (Urbach et al., 2010) (see Table 2 and Table 3).

Manipulation Check

To indicate whether our manipulation was successful and Alan was perceived less human than Philipp Stoecker, a Mann-Whitney-U test was computed, as again, ordinal scales are used and data is not normally distributed. The Mann-Whitey-U test, comparing the perceived humanness and social presence between both groups indicates, that Alan was perceived less human than Philipp Stoecker (W = 3896, p = .003) and a higher social presence was experienced with Philipp Stoecker than with Alan (W = 3847, p = .004). Therefore, our manipulation was successful.

Table 2. Perceived humanness and social presence of Alan and Philipp Stoecker

Construct	**M_{EG}**	**M_{CG}**	**SD_{EG}**	**SD_{CG}**	**Cronbach's alpha**	**CR**	**AVE**	**Shapiro Wilk**
Perceived humanness *(Survey #3)*	3.65	3.09	1.50	1.32	.928	.932	.697	W = .95 $p < .01$
Social presence *(Survey #3)*	4.33	3.69	1.61	1.41	.938	.939	.757	W = .96 $p < .01$

Results regarding hypothesis 1

As our manipulation was successful (**MC**), we continue to test hypothesis 1, answering whether there is a significant difference in evaluation apprehension between individuals presenting their idea to an AI-based computer system and individuals that present their idea to a human. Table 3 shows the descriptive data of the BFNE-R measures of the experimental group and control group and the general BFNE-R value (general fear of being evaluated) of the whole sample. To test hypothesis 1 (**H1**), we computed a Mann-Whitney-U test of the BFNE-R values of the experimental group and the control group. The results show that there is a significant difference (W = 9742, p < .01), concluding that hypothesis 1 (**H1**) is supported. To cross-check the difference, we calculated a Mann-Whitney-U test between the general BFNE-R value of experimental group (M = 3.09) and the control group (M = 3.04), resulting in no significant difference between the groups (W = 6022, p = 0.343). Thus indicating that a distribution between people that generally fear being evaluated and people that don't fear being evaluated is even.

Table 3. Descriptive data and Cronbach's alpha of BFNE-R

Construct	**N**	**M**	**SD**	**Cronbach's alpha**	**CR**	**AVE**	**Shapiro Wilk**
General BFNE-R *(Survey #1)*	228	3.09	0.87	.94	.944	.592	W = .98 $p < .01$
BFNE-R (EG) *(Survey #2)*	117	2.02	0.90	.96	.957	.654	W = .90 $p < .01$
BFNE-R (CG) (Survey #2)	111	2.67	1.03	.96	.962	.682	W = .93 $p < .01$

Results regarding hypothesis 2

In the next step, we analyzed the total sample of our study regarding their general fear of being evaluated (general BFNE-R) in order to test, whether there are significant differences in evaluation apprehension between individuals presenting their idea to an AI-based computer system and individuals that present their idea to a human for people with a low, medium or high value of evaluation apprehension. Figure 3 shows the results of the general BFNE-R of all participants or our study.

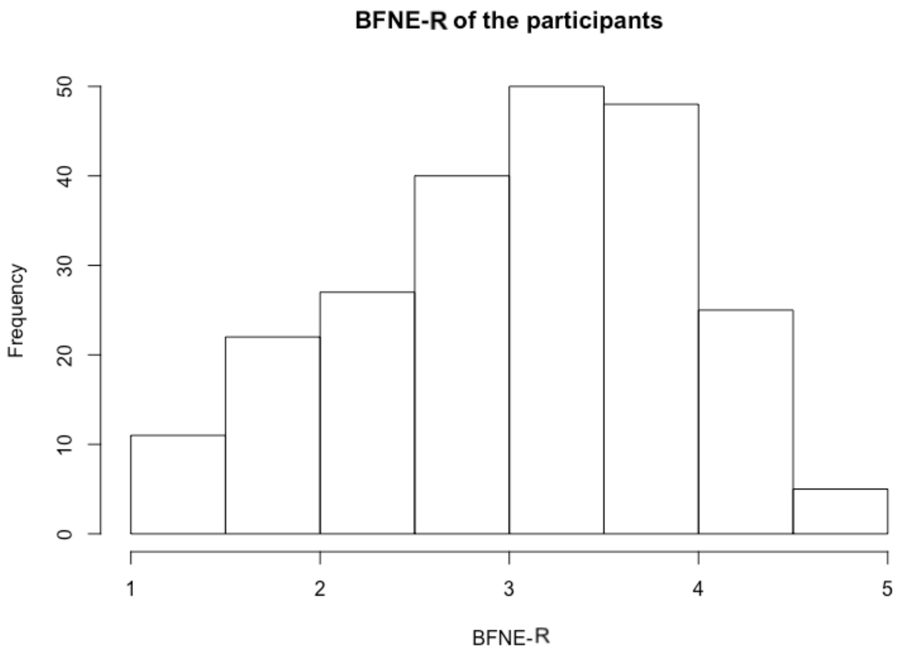


Figure 3. BFNE-R of the participants

The population was divided into three different groups, representing participants with a low general fear of negative evaluation (FNElow), medium general fear of negative evaluation (FNEmed) and high general fear of negative evaluation (FNEhigh). Clustering was performed by means of the R package *Ckmeans.1d.dp*, a procedure for optimal k-means clustering in one dimension (H. Wang & Song, 2011). This resulted in three groups with n = 49 for FNElow (Cluster center = 1.81, Min = 1.00, Max = 2.33), n = 95 for FNEmed (Cluster center = 2.97, Min = 2.42, Max = 3.42) and n = 84 for FNEhigh (cluster center = 3.97, Min = 3.50, Max = 5.0). Table 4 also shows the distribution between the experimental group and the control group according to their general BFNE-R cluster.

Table 4. BFNE-R for each cluster

Cluster	N_{EG}	N_{CG}	**BFNE-R** M_{EG}	**BFNE-R** SD_{EG}	**BFNE-R** M_{CG}	**BFNE-R** SD_{CG}
Low	25	24	1.54	0.57	1.80	0.77
Medium	44	51	2.05	0.80	2.79	0.94
High	48	36	2.24	1.03	3.07	0.98

Of particular interest are the participants with a high value of general fear of being evaluated and those with a medium value and the question of whether there is a significant difference within the experimental group and the control group in their experience evaluation apprehension prior to presenting their idea. To validate these hypotheses, we computed a set of Mann-Whitney-U tests. The results show a significant difference between the experimental group and control group for the cluster FNEhigh (W = 477, $p < .01$) and for the cluster FNEmedium (W = 623, $p < .01$). The cluster FNElow, representing individuals with a low fear of negative evaluation, had no significant difference between the experimental and control condition (W = 241, p = 0.24). Therefore, hypotheses 2a, 2b and 2c are supported.

Results regarding hypothesis 3

In order to test hypothesis 3a-3c, we want to assess whether the general fear of being evaluated is lower than the the evaluation apprehension prior to the idea presentation to Alan (only experimental group) Therefore, we want to investigate whether the fear of being evaluated decreases when presenting an idea to an AI-based system. In doing so, we only look at the participants of the experimental condition and at their general BFNE-R value and their BFNE-R value and whether there is a significant decrease. Individuals with a medium fear of being evaluated that presented their ideas to Alan (n = 44) had an average of 2.99 of general BFNE-R and rated their fear of being evaluated with 2.05, when presenting their idea to Alan, which is a decrease of 31%. Individuals with a high evaluation apprehension that presented their ideas to Alan (n = 48) had an average of 4.01 of general fear of being evaluation and rated their fear of being evaluated with 2.24, when presenting their idea to Alan, which is a decrease of 44%. Individuals with a low evaluation apprehension that presented their ideas to Alan (n = 25) had an average of 1.80 of general fear of being evaluation and rated their fear of being evaluated with 1.54, when presenting their idea to Alan, which is a decrease of 14%. We computed a set of Mann-Whitney-U tests to validate, whether this decrease is significant. The results show a significant difference between the general BFNE-R and the BFNE-R for the people with a medium value of evaluation apprehension. ($W = 345$, $p < .01$) and the people with a high value of evaluation apprehension ($W = 199$, $p < .01$). There is no significant difference for the people with a low value of evaluation apprehension ($W = 408$, $p = 0.07$). Therefore, we can conclude that hypothesis 3a, 3b and 3c are supported.

Discussion

The results indicate that there is a significant difference between the fear of negative evaluation measure of the experimental group and the control group, indicating that the participants express less evaluation apprehension when presenting their idea to Alan. Using an AI-based system for idea evaluation, therefore, causes less feeling of concern or unease about an upcoming evaluation, whereas the evaluation by a human causes higher evaluation apprehension. Exposing an idea to an AI-based system like Alan can make individuals more willing to present their ideas because they are not particularly concerned about a possible negative evaluation by such a system. In contrast, the results show that when humans are involved in evaluating an idea, people tend to feel concerned. Even if that specific human is not known to the participant, the mere presence and perception of a human that is going to evaluate an individually produced idea contributes to the fear of being evaluated.

Furthermore, our manipulation showed that although both characters provide the same information, a system with an avatar of a machine and a description of an AI-based autonomous system is perceived as partially human. However, Alan was perceived as less human than a system with identical interaction but a human representation in terms of its avatar and description. When looking specifically at the individuals with a high or medium value of FNE, the results indicate that an AI-based system like Alan can address the fear of negative evaluation and can be used to evaluate ideas from individuals who would otherwise withhold their ideas because of their fear of being negatively evaluated. Hypothesis 2b and 2c are supported and show that there is a significant difference between the experimental group and the control group in both, the medium and high evaluation apprehension individuals. It seems that individuals with low evaluation apprehension do not care who evaluates their idea, as they generally have no concerns about disclosing an idea for evaluation (hypothesis 2a). This statement is also confirmed by the other exploratory and open-ended questions of the survey, which the participants answered at the end. For example, one participant said that he/she "doesn't care who evaluates my idea and gives feedback as long as it is constructive and helpful feedback." 47% of the participants stated that they either do not care who evaluates their idea (32%) or even prefer an AI-based system for idea evaluation (15%), “if they were to be improved and refined in the future” for example. Other emphasize on the aspects that they wish an AI-based evaluation, “only if constructive criticism is possible”. Furthermore, the aspect that the AI-based system can provide “an unbiased look at your idea from a different angle” is mentioned as a potential benefit.

Our results also show, that the fear of negative evaluation decreased within the medium and high evaluation apprehension group by 31% (medium) and 44% (high), also showing a significant difference when having their ideas evaluated by Alan. This indicates, that an AI-based system can be an appropriate and beneficial tool to address evaluation apprehension. Especially individuals with evaluation apprehension (medium to high) would therefore benefit from such systems. Furthermore, looking at the whole sample across both conditions, we also see a significant decrease of evaluation apprehension ($W = 36752$, $p < .01$), which is not surprising as the fear of being criticized is most present in analog settings in which high social presence is felt and people see, hear, and feel each other (Zhou et al., 2019). In physical settings, in which people work together, more social cues are involved (i.e., mimic and gestures) that can express emotions which furthermore can induce fear of being evaluated. Digital settings lower these effects as people feel a sense of anonymity (Connolly et al., 1990). In addition, hierarchy between the evaluator and the evaluated is an important factor that increases evaluation apprehension, which again could not have occurred due to the nature of our experiment, as Philip is an unknown person who is unrelated to the subject. Our participants had a sense of anonymity while participating in our online experiment which can explain the overall decrease of evaluation apprehension across conditions. This is in line with existing research on anonymity and IT-supported idea generation and evaluation (Zhou et al., 2019).

In addition, 62% of the participants state, that they can imagine having their ideas evaluated by an AI-based system in the future. Participants said, for example, that "AI-based systems could find errors in ideas", "could bring a different perspective to the idea" or "could think well around the corner". However, the participants were also skeptical towards the capabilities of AI and stated that AI is currently not yet capable of evaluating ideas, that “this is a task that should always be done by humans” and that “human minds are superior to AI”.

Implications

Our results lead to several implications for both theory and practice. According to the theory of evaluation apprehension (Henchy & Glass, 1968), individuals fear being evaluated when disclosing ideas to others, which has been shown in our study. This is also the case when it is a person who is neither known nor in a direct relationship with them. Our results are therefore in line with existing theories and studies on evaluation apprehension. Our experiment also offers interesting insights into the question whether theories of human-human interaction (Krämer et al., 2012) and collaboration (Nass et al., 1996; Seeber et al., 2019; Siemon et al., 2018) can be applied to a human-AI interaction without adapting them. In the case of evaluation apprehension, behaviors cannot be transferred directly to human-AI interaction, since humans do not have a feeling of concern when they present their ideas to an AI-based system. Even if they appear human-like and possess human characteristics (e.g. human communication), the theory of evaluation apprehension needs to be refined and more research is needed in order to understand what characteristics of an AI-based system might cause evaluation apprehension in humans. The results also contribute to the understanding of the computers are social actors theory and studies of anthropomorphism in AI-based interaction systems (Nass & Moon, 2000; Qiu & Benbasat, 2009). By using human characteristics (e.g. name and natural language), Alan was perceived as partially human and a certain social presence was felt. Also, further implications are that people with an especially high or moderate fear of criticism feel less to no fear when they reveal an idea to an AI-based system. This implies that an (initial) evaluation by an AI-based system can be useful, especially for such people to pursue possible ideas that would otherwise be lost. Compared to that, individuals with a high FNE have higher feeling of concern and unease when Philipp Stoecker was about to evaluate their ideas.

These findings have far-reaching implications for companies and organizations that could benefit from using AI-based systems in idea and innovation management. Companies should consider using AI-based systems to evaluate initial ideas from employees or customers. Since the fear of evaluation is an enormous factor when it comes to losing creative potential, due to ideas being held back and innovative solutions remaining undiscovered, there should be the possibility to have initial and possibly wild ideas evaluated by an AI-based system in addition to an open, considerate and preferably heterarchical innovation management. This also applies to crowdsourcing platforms, where wild ideas can be evaluated in an initial phase by an AI-based system before being presented to the crowd. Since especially the aspect of public presentation of ideas and possible humiliation by the crowd can be circumvented (Kim et al., 2010). Such systems could be integrated as modular parts in existing idea generation systems. With the help of an active AI-based system, these ideas can then be improved, revised and continuously pursued before they are presented to a potential decision-maker. This could lead to more mature ideas and, potentially lost ideas could be transformed into innovative products and services.

Limitations

Our study comes with a variety of limitations. First, the unrepresentative sample, does not allow a generalization of the results, as most of our 228 participants are between the age of 18 – 29, and mainly students. Furthermore, the nature of the experiment limits a generalization, since both Alan and Philipp Stoecker were simulated. Although the experiment was able to ensure a high internal validity (Karahanna et al., 2018), external validity and realism of the demonstrated phenomenon should be further evaluated.

Other factors could also have played a role in the results, such as the specific characteristics of Alan or Philipp Stoecker. The description of Alan alone, as well as his name, may have influenced evaluation apprehension. The choice of the avatar may also have influenced this phenomenon. Furthermore, the selected photo and the name of Philipp Stoecker may have influenced the control group. Although the photo does not make a particularly friendly or frightening impression, this may be an indicator that a concern of negative evaluation is affected. The chosen gender of both Alan (male name) and Philipp Stoecker (male name) may also have had an influence.

With our experimental design we followed a Wizard of Oz experiment by simulating two idea evaluators, a human and an AI-based system. By simulation both conditions one can argue that no real human was involved in the idea evaluation process which might also limit the transferability of our results to a real scenario with real human actors. Nevertheless, we think that our results can be transferred to humans, since humans interacting via IT-based systems are always only representations of real humans and the abundance of cues that the system can transport determines the perceived humanity and social presence according to the media richness theory (Daft & Lengel, 1986). Our manipulation check showed that the participants felt a high level of social presence when interacting with Philipp Stoecker and that he was also perceived as human.

Besides hierarchy, evaluation apprehension also depends on the expert level of the idea evaluator, as evaluation apprehension increases when individuals were told that they are presenting their idea to experts (DiLiello & Houghton, 2008; Zhou et al., 2019). Even though, we were able to eliminate any aspects of hierarchy in our experiment, by choosing a non-related (and fictional) idea evaluator (Philipp Stoecker), the level of expertise was not mentioned. The perception of whether participants expect a valid, reliable and possible harsh evaluation by either Alan or Philipp Stoecker, might have influenced their level of evaluation apprehension as well. Besides that, participants might also question the fact that an AI-based system is able to conduct a proper evaluation of a human-generated idea. This might have led to an influence on their level of evaluation apprehension.

Conclusion and Outlook

Creative tasks, such as the ability to make discoveries, important for idea generation, or the assessment and further improvement of ideas, important for idea evaluation, are one of the major aims of AI (Boden, 2009). Intelligent systems that can take over such tasks involving creativity represent an enormous potential for organizations and companies. The increase in computing power and the further development of intelligent algorithms have immensely increased the possibilities of AI in recent years. What could initially only be shown in theoretical and philosophical essays (Boden, 1998) has now become reality. Computers actively participate in decision-making processes (Jarrahi, 2018) and assess the novelty of ideas (K. Wang et al., 2019). Within the research field of computational creativity, studies and artifacts show how AI-based systems can take over different creative tasks and thus act as active partners in creative processes (Nass et al., 1996). In our research, we specifically refer to the process of idea evaluation and whether it would be beneficial to use AI to assess and comment on ideas. AI-based system are already used for idea evaluation in various domains, where the evaluation can be compared to that by a human expert (Maher & Fisher, 2012; Varshney et al., 2019; K. Wang et al., 2019). Since there is untapped creative potential caused by the holding back of ideas from individuals who feel fear of negative evaluation, we proposed the research question: To what extent do humans experience evaluation apprehension when AI-based systems evaluate their idea?

Our results show that individuals are not afraid to present their idea to an AI-based computer system, but they sense concern and unease when a person is to evaluate the idea. Using AI-based computer systems for idea evaluation can, therefore, be a way to enhance organizational creativity, foster idea generation and reduce missed ideas and untapped innovative potential.

In the next steps, research and practice should focus on a conceptual elaboration and implementation that uses established methods and procedures of idea evaluation to realize such systems. Although, this depends very much on the domain and industry of the company, established constructs and scales such as those of Dean et al., (2006) can be used to create a basis for AI-based idea evaluation. Even though there are already some approaches and implemented AI-based systems that can automatically evaluate ideas (Maher & Fisher, 2012; Varshney et al., 2019; K. Wang et al., 2019), further research should focus on AI-based idea evaluation and especially evaluate other characteristics of the ideas that go beyond the aspect of novelty. Furthermore, the role of anthropomorphism should continue to be researched, as this has a significant influence on the perception of the AI-based system. If such a system is designed in a very human way, it can be assumed that evaluation apprehension will increase and that people will still be afraid of criticism if they present their ideas to a, particularly human-like system.

Future research is needed to provide further insights on the antecedents and moderators of evaluation apprehension in AI-based idea evaluation, to investigate whether, for example, the feedback itself affects evaluation apprehension and whether it might decrease over time, when people realize that the feedback is more benevolent. In addition, further research is needed to investigate evaluation apprehension in teams or groups where AI-based systems are participating as in our experiment, we investigated a two-way constellation between idea generator and idea evaluator. What impact AI-based systems could have in groups or teams when evaluating ideas of several people or evaluating ideas of a single person openly in front of a group may result in further interesting findings.

In summary, it can be said that AI-based idea evaluation is a way of having one's ideas evaluated and thus being able to improve them. These findings not only contribute to existing theories but also reveal an enormous potential for companies.

References

Aleksander, I. (2017). Partners of humans: A realistic assessment of the role of robots in the foreseeable future. *Journal of Information Technology*, *32*(1), 1–9. https://doi.org/10.1057/s41265-016-0032-4

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, *45*(2), 357–376. https://doi.org/10.1037/0022-3514.45.2.357

Amabile, T. M., & Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior*, *36*, 157–183.

Anderson, J., Rainie, L., & Luchsinger, A. (2018). Artificial intelligence and the future of humans. *Pew Research Center*.

Baer, J. (2014). *Creativity and Divergent Thinking: A Task-Specific Approach*. Psychology Press.

Barczak, G., Lassk, F., & Mulki, J. (2010). Antecedents of Team Creativity: An Examination of Team Emotional Intelligence, Team Trust and Collaborative Culture. *Creativity and Innovation Management*, *19*(4), 332–345. https://doi.org/10.1111/j.1467-8691.2010.00574.x

Barki, H., & Pinsonneault, A. (2001). Small Group Brainstorming and Idea Quality Is Electronic Brainstorming the Most Effective Approach? *Small Group Research*, *32*(2), 158–205. https://doi.org/10.1177/104649640103200203

Basadur, M., Runco, M. A., & Vegaxy, L. A. (2000). Understanding How Creative Thinking Skills, Attitudes and Behaviors Work Together: A Causal Process Model. *The Journal of Creative Behavior*, *34*(2), 77–100. https://doi.org/10.1002/j.2162-6057.2000.tb01203.x

Besold, T. R., Schorlemmer, M., & Smaill, A. (2015). *Computational creativity research: Towards creative machines* (1. Aufl., Bd. 7). Atlantis Press.

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, *103*(1–2), 347–356. https://doi.org/10.1016/S0004-3702(98)00055-1

Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Psychology Press. http://books.google.com/books?hl=en&lr=&id=6Zkm4dz32Y4C&oi=fnd&pg=PR9&dq=info:xYBtAZt-STIJ:scholar.google.com&ots=9cYSkatZjs&sig=qsumsJOVR1IzJj2r-qMSLVK1oDY

Boden, M. A. (2009). Computer Models of Creativity. *AI Magazine*, *30*(3), 23. https://doi.org/10.1609/aimag.v30i3.2254

Brewer, M. B., & Crano, W. D. (2000). Research design and issues of validity. *Handbook of research methods in social and personality psychology*, 3–16.

Camacho, M. L., & Paulus, P. B. (1995). The role of social anxiousness in group brainstorming. *Journal of Personality and Social Psychology*, *68*(6), 1071–1080. https://doi.org/10.1037/0022-3514.68.6.1071

Carleton, R. N., Collimore, K. C., & Asmundson, G. J. G. (2007). Social anxiety and fear of negative evaluation: Construct validity of the BFNE-II. *Journal of Anxiety Disorders*, *21*(1), 131–141. https://doi.org/10.1016/j.janxdis.2006.03.010

Chen, F., Zhang, L., & Latimer, J. (2014). How much has my co-worker contributed? The impact of anonymity and feedback on social loafing in asynchronous virtual collaboration. *International Journal of Information Management*, *34*(5), 652–659. https://doi.org/10.1016/j.ijinfomgt.2014.05.001

Colton, S., López de Mantaras, R., & Stock, O. (2009). Computational Creativity: Coming of Age. *AI Magazine*, *30*(3).

Colton, S., & Wiggins, G. A. (2012). Computational Creativity: The Final Frontier? *Proceedings of the 20th European Conference on Artificial Intelligence*, 21–26. https://doi.org/10.3233/978-1-61499-098-7-21

Connolly, T., Jessup, L. M., & Valacich, J. S. (1990). Effects of Anonymity and Evaluative Tone on Idea Generation in Computer-mediated Groups. *Manage. Sci.*, *36*(6), 689–703. https://doi.org/10.1287/mnsc.36.6.689

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of personality and social psychology*, *9*(3), 245.

Cui, A. S., & Wu, F. (2016). Utilizing customer knowledge in innovation: Antecedents and impact of customer involvement on new product performance. *Journal of the academy of marketing science*, *44*(4), 516–538.

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

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—Why and how. *Knowledge-based systems*, *6*(4), 258–266.

Daugherty, P. R., Wilson, H. J., & Chowdhury, R. (2018). Using Artificial Intelligence to Promote Diversity. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/using-artificial-intelligence-to-promote-diversity/

de Vreede, G.-J., & Briggs, R. O. (2019). A Program of Collaboration Engineering Research and Practice: Contributions, Insights, and Future Directions. *Journal of Management Information Systems*, *36*(1), 74–119. https://doi.org/10.1080/07421222.2018.1550552

Dean, D. L., Hender, J. M., Rodgers, T. L., & Santanen, E. L. (2006). Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation. *Journal of the Association for Information Systems*, *7*(1), 30.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, 1–7.

Dennis, A. R., & Williams, M. L. (2003). Electronic brainstorming: Theory, research, and future directions. In P. B. Paulus & B. A. Nijstad (Hrsg.), *Group creativity: Innovation through collaboration* (S. 160–178). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195147308.003.0008

Diederich, S., Brendel, A., Morana, S., & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems*.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497–509. https://doi.org/10.1037/0022-3514.53.3.497

Diehl, M., & Stroebe, W. (1991). Productivity loss in idea-generating groups: Tracking down the blocking effect. *Journal of Personality and Social Psychology*, *61*(3), 392–403. https://doi.org/10.1037/0022-3514.61.3.392

DiLiello, T. C., & Houghton, J. D. (2008). Creative potential and practised creativity: Identifying untapped creativity in organizations. *Creativity and Innovation Management*, *17*(1), 37–46.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Fink, A., Koschutnig, K., Benedek, M., Reishofer, G., Ischebeck, A., Weiss, E. M., & Ebner, F. (2012). Stimulating creativity via the exposure to other people's ideas. *Human brain mapping*, *33*(11), 2603–2610.

Gabriel, A., Monticolo, D., Camargo, M., & Bourgault, M. (2016). Creativity support systems: A systematic mapping study. *Thinking Skills and Creativity*, *21*, 109–122. https://doi.org/10.1016/j.tsc.2016.05.009

Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-Products and e-Services. *Omega*, *32*(6), 407–424.

Gerber, E. (2009). Using Improvisation to Enhance the Effectiveness of Brainstorming. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 97–104. https://doi.org/10.1145/1518701.1518718

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards Designing Cooperative and Social Conversational Agents for Customer Service. *ICIS 2017 Proceedings*. https://aisel.aisnet.org/icis2017/HCI/Presentations/1

große Deters, F., Tams, S., Johnston, A., & Thatcher, J. (2019). Designing Experimental Studies. *ICIS 2019 Proceedings*. https://aisel.aisnet.org/icis2019/pdws/pdws/8

Henchy, T., & Glass, D. C. (1968). Evaluation apprehension and the social facilitation of dominant and subordinate responses. *Journal of Personality and Social Psychology*, *10*(4), 446–454. https://doi.org/10.1037/h0026814

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, *61*(1), 569–598. https://doi.org/10.1146/annurev.psych.093008.100416

Hoever, I. J., van Knippenberg, D., van Ginkel, W. P., & Barkema, H. G. (2012). Fostering team creativity: Perspective taking as key to unlocking diversity's potential. *The Journal of Applied Psychology*, *97*(5), 982–996. https://doi.org/10.1037/a0029159

Hofmann, S. G., & DiBartolo, P. M. (2014). *Social anxiety: Clinical, developmental, and social perspectives*. Elsevier.

Holtgraves, T., & Han, T.-L. (2007). A procedure for studying online conversational processing using a chat bot. *Behavior research methods*, *39*(1), 156–163.

Huang, Z., Liu, L., Zheng, W., Tan, X., & Zhao, X. (2015). Walking the Straight and Narrow: The Moderating Effect of Evaluation Apprehension on the Relationship between Collectivism and Corruption. *PLoS ONE*, *10*(3). https://doi.org/10.1371/journal.pone.0123859

Hung, S.-Y., Lai, H.-M., Yen, D. C., & Chen, C.-Y. (2017). Exploring the Effects of Team Collaborative Norms and Team Identification on the Quality of Individuals' Knowledge Contribution in Teams. *SIGMIS Database*, *48*(4), 80–106. https://doi.org/10.1145/3158421.3158428

Hwang, A. H.-C., & Won, A. S. (2021). IdeaBot: Investigating Social Facilitation in Human-Machine Team Creativity. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3411764.3445270

Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, *53*(2), 251–267. https://doi.org/10.1006/ijhc.2000.0368

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Jessup, L. M., Connolly, T., & Galegher, J. (1990). The Effects of Anonymity on GDSS Group Process with an Idea-Generating Task. *MIS Quarterly*, *14*(3), 313–321. https://doi.org/10.2307/248893

Kankanhalli, A., Tan, B. C. Y., & Wei, K.-K. (2005). Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation. *MIS Quarterly*, *29*(1), 113–143. https://doi.org/10.2307/25148670

Karahanna, E., Benbasat, I., Bapna, R., & Rai, A. (2018). Editor’s comments: Opportunities and challenges for different types of online experiments. *MIS Quarterly*, *42*(4), iii–x.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*(4), 681–706. https://doi.org/10.1037/0022-3514.65.4.681

Karau, S. J., & Williams, K. D. (2001). Understanding individual motivation in groups: The collective effort model. *Groups at work: Theory and research*, 113–141.

Kim, Y.-H., Chiu, C.-Y., Peng, S., Cai, H., & Tov, W. (2010). Explaining East-West Differences in the Likelihood of Making Favorable Self-Evaluations: The Role of

Evaluation Apprehension and Directness of Expression. *Journal of Cross-Cultural Psychology*, *41*(1), 62–75. https://doi.org/10.1177/0022022109348921

Kraemer, K. L., & King, J. L. (1988). Computer-based Systems for Cooperative Work and Group Decision Making. *ACM Comput. Surv.*, *20*(2), 115–146. https://doi.org/10.1145/46157.46158

Krämer, N. C., von der Pütten, A., & Eimler, S. (2012). Human-Agent and Human-Robot Interaction Theory: Similarities to and Differences from Human-Human Interaction. In M. Zacarias & J. V. de Oliveira (Hrsg.), *Human-Computer Interaction: The Agency Perspective* (S. 215–240). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-25691-2_9

Leary, M. R. (1983). A Brief Version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, *9*(3), 371–375. https://doi.org/10.1177/0146167283093007

Lee, K., Lee, K. Y., & Sheehan, L. (2019). Hey Alexa! A Magic Spell of Social Glue?: Sharing a Smart Voice Assistant Speaker and Its Impact on Users' Perception of Group Harmony. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-019-09975-1

Licuanan, B. F., Dailey, L. R., & Mumford, M. D. (2007). Idea evaluation: Error in evaluating highly original ideas. *The Journal of Creative Behavior*, *41*(1), 1–27. https://doi.org/10.1002/j.2162-6057.2007.tb01279.x

Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, *24*(3), 149–157. https://doi.org/10.1016/j.jsis.2015.08.002

Lubart, T. (2005). How Can Computers Be Partners in the Creative Process: Classification and Commentary on the Special Issue. *Int. J. Hum.-Comput. Stud.*, *63*(4–5), 365–369. https://doi.org/10.1016/j.ijhcs.2005.04.002

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science*, *38*(6), 937–947. https://doi.org/10.1287/mksc.2019.1192

Maedche, A., Morana, S., Schacht, S., Werth, D., & Krumeich, J. (2016). Advanced User Assistance Systems. *Business & Information Systems Engineering*, *58*(5), 367–370.

Maher, M. L., & Fisher, D. H. (2012). Using AI to evaluate creative designs. *DS 73-1 Proceedings of the 2nd International Conference on Design Creativity Volume 1*.

Mamykina, L., Candy, L., & Edmonds, E. (2002). Collaborative Creativity. *Commun. ACM*, *45*(10), 96–99. https://doi.org/10.1145/570907.570940

Massetti, B. (1996). An Empirical Examination of the Value of Creativity Support Systems on Idea Generation. *MIS Quarterly*, *20*(1), 83–97. https://doi.org/10.2307/249543

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12–12. https://doi.org/10.1609/aimag.v27i4.1904

McTear, M., Callejas, Z., & Griol, D. (2016). The conversational interface. *Springer*, *6*(94), 102.

Mumford, M. D., Medeiros, K. E., & Partlow, P. J. (2012). Creative Thinking: Processes, Strategies, and Knowledge. *The Journal of Creative Behavior*, *46*(1), 30–47. https://doi.org/10.1002/jocb.003

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

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

Nunamaker, J. F., Applegate, L. M., & Konsynski, B. R. (1987). Facilitating Group Creativity: Experience with a Group Decision Support System. *Journal of Management Information Systems*, *3*(4), 5–19.

Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, D., & George, J. F. (1991). Electronic meeting systems. *Communications of the ACM*, *34*(7), 40–61.

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in human behavior*, *15*(2), 123–142.

Paulus, P. B., Dzindolet, M., & Kohn, N. W. (2012). Collaborative Creativity—Group Creativity and Team Innovation. In *Handbook of Organizational Creativity* (S. 327–357). Academic Press. http://www.sciencedirect.com/science/article/pii/B9780123747143000148

Pisano, G. P. (2015). You need an innovation strategy. *Harvard Business Review*, *93*(6), 44–54.

Porra, J., Lacity, M., & Parks, M. S. (2019). "Can Computer Based Human-Likeness Endanger Humanness?" – A Philosophical and Ethical Perspective on Digital Assistants Expressing Feelings They Can't Have". *Information Systems Frontiers*. https://doi.org/10.1007/s10796-019-09969-z

Power, D. J., & Sharda, R. (2009). Decision Support Systems. In *Springer Handbook of Automation* (S. 1539–1548). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78831-7_87

Qiu, L., & Benbasat, I. (2009). Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. *Journal of Management Information Systems*, *25*(4), 145–182. https://doi.org/10.2753/MIS0742-1222250405

Rao, V. S. C. (1992). Group Support Systems in Organizations: The Potential Effects of Anonymity. *Journal of Information Technology Management*, *3*(2), 29.

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

Reichenberger, J., Schwarz, M., König, D., Wilhelm, F. H., Voderholzer, U., Hillert, A., & Blechert, J. (2015). Angst vor negativer sozialer Bewertung: Übersetzung und Validierung der Furcht vor negativer Evaluation–Kurzskala (FNE-K). *Diagnostica*, *62*(3), 169–181. https://doi.org/10.1026/0012-1924/a000148

Rhodes, M. (1961). An Analysis of Creativity. *The Phi Delta Kappan*, *42*(7), 305–310.

Rickenberg, R., & Reeves, B. (2000). The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 49–56. https://doi.org/10.1145/332040.332406

Riek, L. D. (2012). Wizard of oz studies in hri: A systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, *1*(1), 119–136.

Runco, M. A. (2003). Idea evaluation, divergent thinking, and creativity. In *Critical creative processes* (S. 69–94). Hampton Press.

Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, *24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* ((4th Ed.)). Malaysia; Pearson Education Limited,.

Rzepka, C., & Berger, B. (2018). User Interaction with AI-enabled Systems: A Systematic Review of IS Research. *ICIS 2018 Proceedings*. https://aisel.aisnet.org/icis2018/general/Presentations/7

Santanen, E. L., Briggs, R. O., & Vreede, G.-J. D. (2004). Causal Relationships in Creative Problem Solving: Comparing Facilitation Interventions for Ideation. *J. Manage. Inf. Syst.*, *20*(4), 167–198.

Sawyer, R. K., & DeZutter, S. (2009). Distributed creativity: How collective creations emerge from collaboration. *Psychology of Aesthetics, Creativity, and the Arts*, *3*(2), 81–92. https://doi.org/10.1037/a0013282

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2019). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 103174. https://doi.org/10.1016/j.im.2019.103174

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry, P. B. (2020). Collaborating with technology-based autonomous agents. *Internet Research*.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2000). *Research methods in psychology*. McGraw-Hill.

Shepherd, M. M., Briggs, R. O., Reinig, B. A., Yen, J., & Nunamaker, J. F. (1995). Invoking Social Comparison to Improve Electronic Brainstorming: Beyond Anonymity. *J. Manage. Inf. Syst.*, *12*(3), 155–170. https://doi.org/10.1080/07421222.1995.11518095

Siemon, D., Eckardt, L., & Robra-Bissantz, S. (2015). Tracking Down the Negative Group Creativity Effects with the Help of an Artificial Intelligence-Like Support System. *2015 48th Hawaii International Conference on System Sciences (HICSS)*, 236–243. https://doi.org/10.1109/HICSS.2015.37

Siemon, D., Strohmann, T., & Michalke, S. (2022). Creative Potential Through Artificial Intelligence: Recommendations for Improving Corporate and Entrepreneurial Innovation Activities. *Communications of the Association for Information Systems*.

Siemon, D., Strohmann, T., & Robra-Bissantz, S. (2018). The Virtual Collaborator—A Definition and Research Agenda. *International Journal of E-Collaboration (IJeC)*, *14*(4), 24–43. https://doi.org/10.4018/IJeC.2018100102

Simons, L. S., & Turner, C. W. (1976). Evaluation apprehension, hypothesis awareness, and the weapons effect. *Aggressive Behavior*, *2*(1), 77–87. https://doi.org/10.1002/1098-2337(1976)2:1<77::AID-AB2480020108>3.0.CO;2-A

Somech, A., & Drach-Zahavy, A. (2013). Translating Team Creativity to Innovation Implementation The Role of Team Composition and Climate for Innovation. *Journal of Management*, *39*(3), 684–708. https://doi.org/10.1177/0149206310394187

Sosik, J. J., Kahai, S. S., & Avolio, B. J. (1998). Transformational leadership and dimensions of creativity: Motivating idea generation in computer-mediated groups. *Creativity Research Journal*, *11*(2), 111–121.

Spinella, L. (2018). *Intelligent Virtual Assistants 101*. Interactions Resource Center. https://resources.interactions.com/library/intelligent-virtual-assistants-101/

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-computer interaction*, *11*(2), 97–124.

Stieglitz, S., Brachten, F., & Kissmer, T. (2018). Defining Bots in an Enterprise Context. *International Conference on Information Systems*.

Strohmann, T., Siemon, D., & Robra-Bissantz, S. (2017). brAInstorm: Intelligent Assistance in Group Idea Generation. *International Conference on Design Science Research in Information System and Technology*, 457–461.

Tasa, K., Taggar, S., & Seijts, G. H. (2007). The development of collective efficacy in teams: A multilevel and longitudinal perspective. *The Journal of Applied Psychology*, *92*(1), 17–27. https://doi.org/10.1037/0021-9010.92.1.17

Thomas, J., Hall, J. A., Miller, F. D., Dewhirst, J. R., Fine, G. A., Taylor, M., & Rosnow, R. L. (1979). Evaluation apprehension, social desirability, and the interpretation of test correlations. *Social Behavior and Personality*, *7*(2), 193–198. https://doi.org/10.2224/sbp.1979.7.2.193

Urbach, N., Ahlemann, F., & others. (2010). Structural equation modeling in information systems research using partial least squares. *Journal of Information technology theory and application*, *11*(2), 5–40.

Varshney, L. R., Pinel, F., Varshney, K., Bhattacharjya, D., Schörgendorfer, A., & Chee, Y.-M. (2019). A big data approach to computational creativity: The curious case of Chef Watson. *IBM Journal of Research and Development*, *63*(1), 7–1.

Verhagen, T., Van Nes, J., Feldberg, F., & Van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, *19*(3), 529–545.

Wang, H., & Song, M. (2011). Ckmeans. 1d. Dp: Optimal k-means clustering in one dimension by dynamic programming. *The R journal*, *3*(2), 29.

Wang, K., Dong, B., & Ma, J. (2019). *Towards Computational Assessment of Idea Novelty* (SSRN Scholarly Paper ID 3393611). Social Science Research Network. https://papers.ssrn.com/abstract=3393611

Wang, K., & Nickerson, J. V. (2017). A literature review on individual creativity support systems. *Computers in Human Behavior*, *74*, 139–151. https://doi.org/10.1016/j.chb.2017.04.035

Wang, X., Schneider, C., & Valacich, J. S. (2015). Enhancing creativity in group collaboration: How performance targets and feedback shape perceptions and idea generation performance. *Computers in Human Behavior*, *42*, 187–195. https://doi.org/10.1016/j.chb.2014.02.017

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, *33*(4), 448–457. https://doi.org/10.1037/h0027806

Wilson, C. (2013). *Brainstorming and beyond: A user-centered design method*. Newnes.

Woods, S., Dautenhahn, K., & Kaouri, C. (2005). Is someone watching me? - Consideration of social facilitation effects in human-robot interaction experiments. *2005 International Symposium on Computational Intelligence in Robotics and Automation*, 53–60. https://doi.org/10.1109/CIRA.2005.1554254

Wright, M. F. (2013). The relationship between young adults' beliefs about anonymity and subsequent cyber aggression. *Cyberpsychology, Behavior, and Social Networking*, *16*(12), 858–862.

Wu, H., & Leung, S.-O. (2017). Can Likert Scales be Treated as Interval Scales?—A Simulation Study. *Journal of Social Service Research*, *43*(4), 527–532. https://doi.org/10.1080/01488376.2017.1329775

Yan, J. (Kevin), Leidner, D. E., & Benbya, H. (2018). Differential Innovativeness Outcomes of User and Employee Participation in an Online User Innovation Community. *Journal of Management Information Systems*, *35*(3), 900–933. https://doi.org/10.1080/07421222.2018.1481669

Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, *16*(3), 417–434. https://doi.org/10.1007/s10796-012-9350-4

Zhou, X., Delidabieke, B., Zhai, H., Zeng, H., & Cao, X. (2019). Exposure to Ideas, Evaluation Apprehension, and Incubation Intervals in Collaborative Idea Generation. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01459

Appendix

Table 5. The written evaluation by Alan and Philipp Stoecker

Evaluation by Alan and Philipp Stoecker
I like your idea very much and at first sight it seems to me to be new and very innovative. As far as I know, such a concept does not yet exist and therefore it should definitely be pursued further. However, if you take a closer look, I think that the feasibility of your concept is not easy and above all it is difficult to realize economically. You should therefore think again carefully whether certain things could be tackled more simply and thus also more economically. This should be elaborated further. I also think that your idea cannot be generalized, so a specification would be useful. Despite its uniqueness, the idea still has some weaknesses that should be worked out. I hope I could help you with my feedback.

Table 6. Measures

Brief Fear of Negative Evaluation-Revised (Carleton et al., 2007) - (General) **(1=totally disagree, 5=totally agree)**
1. I worry about what other people will think of me even when I know it doesn't make any difference.
2. It bothers me when people form an unfavourable impression of me.
3. I am frequently afraid of other people noticing my shortcomings.
4. I worry about what kind of impression I make on people.
5. I am afraid that others will not approve of me.
6. I am afraid that other people will find fault with me.
7. I am concerned about other people's opinions of me.
8. When I am talking to someone, I worry about what they may be thinking about me.
9. I am usually worried about what kind of impression I make.
10. If I know someone is judging me, it tends to bother me.
11. Sometimes I think I am too concerned with what other people think of me.
12. I often worry that I will say or do wrong things.
Brief Fear of Negative Evaluation-Revised (Carleton et al., 2007) - (Experienced) **(1=totally disagree, 5=totally agree)**
1. I worry about what Alan/Philipp Stöcker will think of me even when I know it doesn't make any difference.
2. It bothers me when Alan/Philipp Stöcker form an unfavourable impression of me.

3. I am afraid of Alan/Philipp Stöcker noticing my shortcomings.
4. I worry about what kind of impression I make on Alan/Philipp Stöcker.
5. I am afraid that Alan/Philipp Stöcker will not approve of me.
6. I am afraid that Alan/Philipp Stöcker will find fault with me.
7. I am concerned about Alan's/Philipp Stöcker's opinions of me.
8. When I am presenting my idea to Alan/Philipp Stöcker, I worry about what he may think about me.
9. I am worried about what kind of impression I make, when presenting my idea to Alan/Philipp Stöcker.
10. If I know that Alan/Philipp Stöcker is judging me, it tends to bother me.
11. I think I am too concerned with what Alan/Philipp Stöcker thinks of me.
12. I worry that I will say or do wrong things in front of Alan/Philipp Stöcker.
Perceived humanness (Holtgraves and Han 2007) ***(1=totally agree, 7=totally disagree)***
How human-like do you perceive Alan/Philipp Stoecker?
How skilled do you perceive Alan/Philipp Stoecker?
How thoughtful do you perceive Alan/Philipp Stoecker?
How polite do you perceive Alan/Philipp Stoecker?
How responsive do you perceive Alan/Philipp Stoecker?
How engaging do you perceive Alan/Philipp Stoecker?
Social presence (Gefen and Straub 1997) ***(1=totally agree, 7=totally disagree)***
I felt a sense of human contact with Alan/Philipp Stoecker.
I felt a sense of personalness with Alan/Philipp Stoecker.
I felt a sense of sociability with Alan/Philipp Stoecker.
I felt a sense of human warmth with Alan/Philipp Stoecker.
I felt sense of human sensitivity with Alan/Philipp Stoecker.
Open questions
What is your opinion about Alan's/Philipp Stoecker's evaluation?
Do you have any other comments or ideas?