6-1-2023

TOWARDS HUMAN-AI-COLLABORATION IN BRAINSTORMING: EMPIRICAL INSIGHTS INTO THE PERCEPTION OF WORKING WITH A GENERATIVE AI

Lucas Memmert

Navid Tavanapour

TOWARDS HUMAN-AI-COLLABORATION IN BRAINSTORMING: EMPIRICAL INSIGHTS INTO THE PERCEPTION OF WORKING WITH A GENERATIVE AI

Research Paper

Lucas Memmert, Universität Hamburg, Germany, lucas.memmert@uni-hamburg.de.

Navid Tavanapour, Universität Hamburg, Germany, navid.tavanapour@uni-hamburg.de.

Abstract

Groups of humans or crowds can be remarkable when coming up with ideas. However, not everyone has a group of humans at their disposal to brainstorm. With recent advances in AI, however, generative large language models (LLM) might be capable of contributing ideas in a brainstorming session, turning individual work of a human into joint work of human and AI. It is, however, unclear, how group effects known from human brainstorming groups transfer to such a human-AI setting. In our mixed-method study (qualitative emphasis) with 24 participants, we investigate how a human brainstorms together with the generative LLM 'GPT-3', and how they perceived their experience. Our results highlight known effects like cognitive stimulation but also a risk of free riding. We thereby contribute to the understanding of how generative AI, which is becoming broadly available, can be used to address the challenge of human-AI collaboration for solving open-ended problems.

Keywords: Human AI collaboration, Brainstorming, GPT-3, Cognitive stimulation, Free Riding.

1 Introduction

Solving problems is a key activity for organizations (Nickerson and Zenger, 2004). Brainstorming is a popular technique for groups of humans to develop new ideas on a topic or problem (Osborn, 1953; Maaravi et al., 2021; Hickey et al., 2003). However, not everyone faced with a problem requiring creative ideas has a group of humans cost-effectively available to brainstorm ideas for their problem or is allowed to share the problem with the world. With advances of machines the question arises whether AI can "augment such human problem-solving performance by […] discovering valuable candidate solutions" (Krogh, 2018) and collaboration of humans and AI systems could be achieved to solve complex problems (Dellermann et al., 2019). The idea to support individual humans on content-level in creative tasks is not new (Wang and Nickerson, 2017), several systems were proposed, e.g., by offering stimuli based on association dictionaries (Althuizen and Reichel, 2016), or by querying social media for inspiration (Siemon et al., 2015). However, such systems did not generate and contribute complete ideas like another human. Traditionally, creativity is described as a human strengths (Dellermann et al., 2019; Lichtenthaler, 2018; Krogh, 2018). With recent advances in generative language models (GLMs), however, human-AI collaboration for creative tasks might be feasible (Gero et al., 2022). Similar to us, Di Fede et al. (2022) suggested exploring using GLMs to support humans in generating ideas.
From group brainstorming literature, several effects are known to influence group performance when generating ideas (Pinsonneault et al., 1999). Examples include group members being stimulated by the exposure to other group members' ideas to come up with additional ideas (*cognitive stimulation*) enhancing performance, or, group members reducing their effort as soon as they are part of group (*free riding),* reducing group performance (Pinsonneault et al., 1999). If a human works together with an AI

system that contributes ideas, individual work of a human is transformed into collaborative work of a human and an AI system. While it is known that humans tend to anthropomorphize AI systems (Salles et al., 2020), it is unclear if typical effects known from human brainstorming groups occur in human-AI brainstorming groups too. However, this is important to understand to design effective human-AI collaboration. Therefore, we seek to investigate how individual humans work with and perceive working with such a novel AI system to answer the following research question: *How do individual humans perceive working together with a generative AI system in a brainstorming activity, particularly with regard to known group effects?*

In our exploratory, mixed-method study (qualitative emphasis), participants individually brainstorm with a generative AI and reflect on their experience afterward via a qualitative survey. We show that signs of group effects (e.g., *cognitive stimulation* and *free riding, Pinsonneault et al., 1999*) known from human brainstorming sessions partially occur in such human-AI groups as well. More broadly, we show that an off-the-shelve, non-configured or adjusted GLM (GPT-3) is capable of providing AI suggestions perceived to be helpful and inspirational. We thereby take a first step towards addressing the broader research challenge of designing human-AI collaboration to solve (complex) problems (Dellermann et al., 2019; Akata et al., 2020). More specifically, we explore the team dynamics of humans working with AI systems (Makarius et al., 2020), which could inform designing effective human-AI collaboration (Seeber et al., 2020), with the AI system taking the role of an expert or peer as opposed to a facilitator (Seeber et al., 2020) in an open-ended problem-solving task as opposed to closed-ended decision problem (Krogh, 2018). With powerful GLMs becoming broadly available and embedded into many applications, we believe our research to be timely.

2 Background

We briefly highlight *group effects* known from all-human brainstorming groups. We relate our approach to research around supporting brainstorming through (AI) technology and argue that a new type of AI, generative AI might now enable content-level collaboration through its generative capability.

2.1 Brainstorming

'Brainstorming' is a creative technique proposed by Osborn (1953) for groups of humans to generate ideas on a creative problem. Osborn (1953) suggested four rules for the group members to follow: (1) delayed judgment, (2) encouragement of wild ideas, (3) focus on quantity of ideas, and (4) encouragement of improving and combining ideas. Since then, lots of research has been conducted on performing brainstorming effectively. One stream of research investigated how different group settings affected brainstorming performance. Brainstorming could, e.g., be performed face-to-face or electronically (electronic brainstorming, EBS) with verbal or non-verbal expression of ideas, as a group ("real group") or individually with pooling of ideas ("nominal group"), or with ideas contributed anonymously or non-anonymously (Pinsonneault et al., 1999; Cooper et al., 1998; Dennis and Williams, 2005). Depending on the setting, different effects influence the group's brainstorming performance. Providing a comprehensive overview of prior research, Pinsonneault et al. (1999) listed 16 effects: six process gains (i.e., performance-enhancing effects) and ten process losses (i.e., performance-reducing effects). We highlight some effects to provide context for our study.

The process gain of *cognitive stimulation*, i.e., the stimulation of new ideas through the utterances of other group members, is only present in *real groups*, as members of nominal groups do not see utterances of other team members (Pinsonneault et al., 1999). Exposure to other's ideas, e.g., presented via an audio tape (Dugosh et al., 2000) or through a confederate (Paulus et al., 2013) can improve the brainstorming performance, with exposure to more ideas and higher attention paid having a positive effect (Leggett Dugosh and Paulus, 2005). Besides stimulation through others' ideas, cognitive priming, i.e., subconscious stimulation of the working memory, was also shown to be effective in improving brainstorming performance (Dennis et al., 2013). During verbal brainstorming, the process loss *production blocking* might occur (Diehl and Stroebe, 1987), which refers to group members "being unable to express ideas as they occur", e.g., because another group member is speaking and they forget their idea (Pinsonneault et al., 1999). One way to reduce or prevent production blocking is by performing EBS, where humans can type their ideas without waiting for others to finish. Even for EBS with *real groups* different settings have been investigated, such as comparing anonymous and non-anonymous idea contribution (e.g., Cooper et al., 1998; Pinsonneault et al., 1999). While the former can prevent the process loss of *Evaluation apprehension*, i.e., "members fear expressing ideas because of potential retaliation" (Pinsonneault et al., 1999), there also is a risk of free-riding, i.e., "members might limit their efforts and contributions by relying on others to accomplish the task" (Pinsonneault et al., 1999), leading to performance loss. These are only some of the effects that were studied to understand performance differences. However, the focus traditionally was on human groups. It is unclear if those process gains and losses also occur in our proposed setting of a human brainstorming with a generative AI system.

2.2 Human-AI collaboration in brainstorming

Humans and AI systems working together was discussed under different labels such as human-AI collaboration or hybrid intelligence (e.g., Dellermann et al., 2019; Akata et al., 2020), with the goal to achieve "superior results to those each of them could have accomplished separately" (Dellermann et al., 2019). While recent advances in AI enabled machines to outperform humans in several areas, combining humans and AI to solve complex problems might be the next challenge (Dellermann et al., 2019). One aspect discussed in this context was the role of the AI. Bittner et al. (2019) developed a taxonomy for conversational agents, which we, similar to Siemon et al. (2020), interpret more broadly for human-AI collaboration. They differentiate three roles: facilitators "guide users to reach a certain goal or execute a task", peers "aim to merge into a human group […] or become a sparring partner for an individual", and experts "that have certain skills or fields of expertise that differ from those of their human teammates, but mostly act rather reactively upon request". Concerning ideation or brainstorming, there has been much research on a facilitative level (Wang and Nickerson, 2017). Earlier research on EBS investigated the effects of using different facets of technology to conduct brainstorming sessions. More recently, AI-enabled tools have been used to guide participants through brainstorming sessions, e.g., conversational agents that react to user inputs or record user ideas (Tavanapour et al., 2020). However, such support typically occurs on a meta-level, i.e., the system does not contribute ideas to solve the problem at hand (i.e., process guidance independently of the specific problem). There was less focus on creating AI systems that support brainstorming on content-level. The nature of the task might explain this. Developing a system that implements procedures to guide humans through the brainstorming process is one thing. However, building a system capable of contributing creative ideas like a human might be difficult. Particularly so, as a priori, it is unknown which brainstorming questions the users will want to use the system for and the system would need to offer suggestions for any question the user might want to brainstorm. While on a facilitative-, meta-level the same procedure might be used across any brainstorming task, suggestions would have to be different for every brainstorming question.

Although it was shown that offering creative stimuli can significantly improve brainstorming performance (cognitive stimulation) only few systems attempt brainstorming content-level support (Siangliulue et al., 2015b). Althuizen and Reichel's (2016) prototype used a word associations dictionary from which they pre-selected a subset of words related to their specific problem to offer inspiration. Siemon et al. (2015) developed a prototype to aid humans in brainstorming by showing related social media content. They transformed user's ideas into search queries for web and social media platforms and showed the queried content generated by others to offer inspiration. While these approaches affect performance processes, they cannot generate (new) ideas in a comparable manner to a human team member. Creativity was traditionally seen as a human strength (Dellermann et al., 2019; Lichtenthaler, 2018), and AI might be considered constrained due to having to draw on historical data (Krogh, 2018). We suggest exploring state-of-the-art GLMs, which might be able to contribute to creative tasks (e.g., Gero et al., 2022; Memmert and Bittner, 2022). With their "Idea Machine" Di Fede et al. (2022) proposed a similar prototype idea to the idea presented in this paper, based on GPT-3, however, they have not (yet) reported empirical data on its effects. We believe this to be a new approach of identifying useful stimuli, an "important research topic" (Wang and Nickerson, 2017), without having to rely on "building or employing a semantic network" as suggested by Wang and Nickerson (2017).

In supporting a human more holistically on content-level, such a system might be considered to take the role of a *peer* or *expert* (Bittner et al., 2019), or a *creator* in “reflect[ing] creative skills, such as finding many possible solutions or searching for new ideas and developments” (Siemon, 2022).

2.3 Content-level support through generative AI

As described above, supporting open-ended, creative problems like brainstorming tasks with AI might be difficult (Lee et al., 2022; Krogh, 2018), as one cannot anticipate all the possible questions to be brainstormed. This makes developing a pool of ideas or suggestions in advance – from which AI system could choose – impossible. Instead, the AI would need to be flexible to address questions as they arise. With traditional natural language processing (NLP) techniques, models were trained for a specific task (e.g., translation, entity recognition). With GPT-3, a generative large language model was developed that outperforms traditional models in many of those tasks but does not require task-specific training (Brown et al., 2020). Even more so, GPT-3 was trained on a broad data set to predict the next word given a certain input. Thus, it can not only perform typical NLP tasks but continue any user input (e.g., Brown et al., 2020; Gero et al., 2022; Lee et al., 2022). Based on this prior research, we argue that such flexible GLMs might allow collaboration with humans in new ways. They neither require pre-selecting a subset of words from a dictionary (Althuizen and Reichel, 2016), limiting practical usability, nor only show social media search results (Siemon et al., 2015). Instead, actual ideas generated on the spot are offered as content-level support. Research on using GLMs is timely as powerful GLMs are becoming available as cloud-hosted services allowing cost-effective usage even for smaller organizations.

In summary, GLMs like GPT-3, an AI-based technology, which might be capable of addressing open-ended problems without requiring to gather training data first, becomes broadly available and embedded into many products. This could enable human-AI collaboration on the open-ended problem of brainstorming, turning individual human work into collaborative work of human and AI. As part of this collaboration the AI would not act as a facilitator but in the role of a peer/expert or creator, contributing ideas similar to a human. However, it is unclear, how group processes affecting performance (e.g., cognitive stimulation, free riding) observed in human groups transfer to this new human-AI setting.

3 Method

Our research goal is to gain a rich understanding of the perception of humans brainstorming with our generative AI system. Given the capabilities of state-of-the-art, generative AI systems might open new ways of human-AI collaboration, we argue it might be difficult for participants to imagine how they would (hypothetically) work with such as system. We thus decided to implement a web application to have participants work together with such as system. To gain a rich understanding of the perception of working with the AI, we employed a qualitative approach (Döring and Bortz, 2016, p. 26), more specifically, we conducted a qualitative survey after the brainstorming session. Qualitative surveys are sometimes referred to as “semi-standardized questionnaires” (Döring and Bortz, 2016, p. 403) as only questions are included but no options for answers. Given the novelty of the technology enabling the system to provide ideas similar to a human, this method allowed participants to describe their experience in their own words (Döring and Bortz, 2016, p. 404). Additionally to the subjective experience reported in the survey, we used the system log data of participants working with the web application to get an account of actual usage. This mixed-method-approach (figure 1) allows us to get a more complete picture of the phenomenon (Venkatesh et al., 2013), allowing us to compare different perspectives drawn from different sources of data (Creswell, 2014). We put more emphasize on the qualitative aspect (Creswell, 2014) as our focus is to gain a rich understanding of the perception of working with the AI.

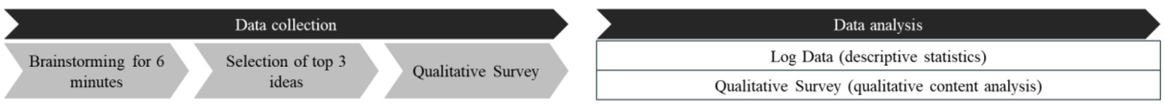

Figure 1. Study method overview

3.1 Data collection

3.1.1 Environment

For data collection we used a common approach from brainstorming literature (e.g., Siemon et al., 2015; Siangliulue et al., 2015a). We set up a brainstorming web application (figure 2) with a task introduction, a timer, the brainstorming question and the text field for users to add their ideas. Submitted ideas were displayed below and could be edited and deleted. Users could request and see AI suggestions next to their own ideas and accepted (i.e., copy) them if they liked. To produce the AI suggestions we used the state-of-the-art, commercial, generative, GLM 'GPT-3' at default settings[1]. We did not make any changes to the model including task (brainstorming) or topic-specific (brainstorming regarding sustainability) training. To produce the suggestions we developed a prompt template (table 1), which we populated with up to three randomly sampled participant ideas (i.e., not shared between participants). These ideas act as 'demonstrations' to condition the GLM's output ('few shot learning', Brown et al., 2020), without requiring model re-training (i.e., adjusting model weights). We included human ideas into the prompt, as one brainstorming rule is to build upon other's ideas (Osborn, 1953) and the AI could thus "consider" the ideas generated by the human. For developing the template, we employed prompt engineering techniques, particularly itemization (Mishra et al.). We developed a structure consisting of a *task*, the brainstorming *question*, and the start of an answer. To implement the itemization technique we indicated the number of items we expected GPT-3's output to contain in the task (number of user suggestions + three new suggestions) and used an enumeration to indicate a start of a list (of ideas). We display the post-processed output (removed enumeration, split into individual suggestions).

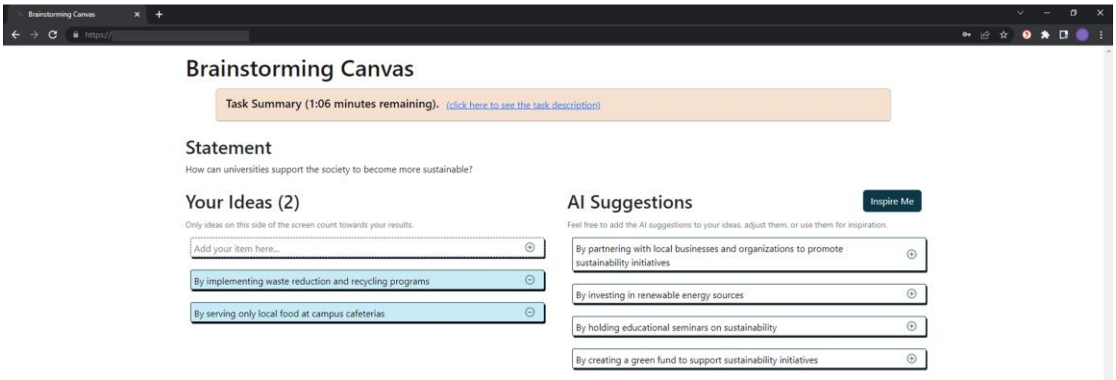

Figure 2. Screenshot of the brainstorming web app (with exemplary ideas)

Prompt template	**Prompt template populated with three user ideas**
Task: Provide 3 ideas to the question below. Question: How can universities support the society to become more sustainable? Answer: 1. by	Task: Provide 6 ideas to the question below. Question: How can universities support the society to become more sustainable? Answer: 1. By conducting research on sustainability 2. By supporting and implementing sustainable practices on campus 3. By investing in renewable energy sources 4.

Table 1. (Empty) prompt template (left) and prompt template populated with three randomly sampled ideas, i.e., 'demonstrations', of one user (right)

[1] The "temperature" parameter was set to 0.9 as recommended by the GPT-3 documentation for "creative applications" (https://beta.openai.com/docs/api-reference/completions/create#completions/create-temperature)

According to the described use case users will be aware that suggestions do not originate from other humans, we thus clearly labeled them as AI suggestions to produce a realistic scenario. During the session users need to actively request AI suggestions as it was shown that providing examples on-demand is more effective than automatically providing examples (Siangliulue et al., 2015b).

3.1.2 Participants and procedure

In total 25 participants from three university courses completed the exercise with a completion time of about 25 minutes. We asked users to brainstorm as many ideas as possible (brainstorming rule according to Osborn, 1953) for six minutes. Universities are a frequent topic of brainstorming studies (e.g., Dennis and Valacich, 1993; Kohn and Smith, 2011; Paulus et al., 2013; Baruah and Paulus, 2016), so are societal problems (e.g., Huber et al., 2019; Zhu et al., 2020). We combined the two themes into the question: “How can universities support the society to become more sustainable?”. It was explained to participants that they could request AI suggestions during the session and could use those suggestions as they liked. As we sought to investigate humans working with the AI system and in brainstorming building ideas upon other’s ideas is encouraged (brainstorming rule according to Osborn, 1953), we asked participants to request AI suggestions at least once. After the session, participants were asked to select their top three ideas (Siangliulue et al., 2015a) and fill in the qualitative survey (Döring and Bortz, 2016).

3.1.3 Measures

Similar to Siemon et al. (2015) we use two sources of data: recoded actual usage data (i.e., log data) and participants’ reported perception. This triangulation allowed us to get a more in-depth understanding of the usage behavior. For the usage data, we recorded all interactions of the participants with the system with timestamps, including start time of the brainstorming session, submission of every idea, every change to ideas, every suggestion request, and every copying of AI suggestions. For the perceived experience data, we developed a qualitative survey (Döring and Bortz, 2016) consisting of 18, mostly open-ended questions regarding the thought process and experience during the session. We developed the questions based on the group effects to potentially occur as summarized by Pinsonneault et al. (1999) and introduced above.

3.2 Data analysis

For the two data sources (i.e., usage data, survey responses) we used different data analysis techniques. For the usage data, we performed descriptive statistical analysis to get a foundational understanding of how the humans worked with the AI system during the brainstorming session. Due to a technical error, some participants were able to submit ideas after the time allocated for the brainstorming session passed. For reporting the results, we filtered out those ideas if not indicated otherwise. To get a more in-depth understanding, we then performed a qualitative content analysis of the open-ended survey responses following Mayring (2014). As our research goal was to understand the group effects within this human-AI setting, we used a deductive approach in which we used the group effects, i.e., the process gains and process losses (e.g., cognitive stimulation, free riding, observational learning) in brainstorming reported in Pinsonneault et al. (1999) as a foundation. We analyzed the responses for indications of such effects and coded them accordingly in MaxQDA (VERBI Software, 2019). We used the descriptions of the group effects by Pinsonneault et al. (1999, Figure 2) as guidance. Thereby we ensured a theoretical grounding with clear reference to our research question as suggested by Mayring (2014). Additionally, we inductively coded the participant’s assessments of the AI suggestions to better understand their perception and what they liked or disliked about the suggestions, and participant’s comments with regard to working with the tool to identify potential improvements for follow-up studies. This additional inductive coding allowed us to gain a more complete understanding of participants’ experiences without “preconceptions of the researcher” (Mayring, 2014, p. 79).

4 Results

4.1 Overview: working with the AI system

We had 25 study participants completing the exercise in three university seminars on design science research, data-driven solutions for smart cities, and human-centered AI. Due to insufficient completion of the survey one participant was excluded. The remaining 24 participants (female=6, male=18) had an average age of 26.8 years and were bachelor or master students in study programs of the informatics department (except for one biology student), the majority in informatics (n=12) and information systems (n=7). They generated 98 ideas (mean=4.1) during the six-minute brainstorming sessions. As our goal was to investigate the participants' perception of working with the AI systems, we asked them to request AI suggestions at least once. However, most participants made multiple requests for AI suggestions (mean=4.5). Participants made 109 requests, received 326 AI suggestions, and accepted (i.e., copied) 156 AI suggestions (mean=6.5). An overview regarding the timelines for each brainstorming session is shown in figure 3. On average, participants added 1.1 ideas prior to making their first request 56.8 seconds into the session. Across participants, splitting the session in 1-minute intervals, participants jointly made 19, 19, 19, 16, 14, and 22 requests respectively. Examples of AI suggestions include (original spelling): "By increasing the number of vegan/vegetarian options in cafeterias", "By Give awards to students/staff/teams who have contributed to making the university more sustainable", and "By have a "sustainability week" where the school shuts down all unnecessary lights and appliances".

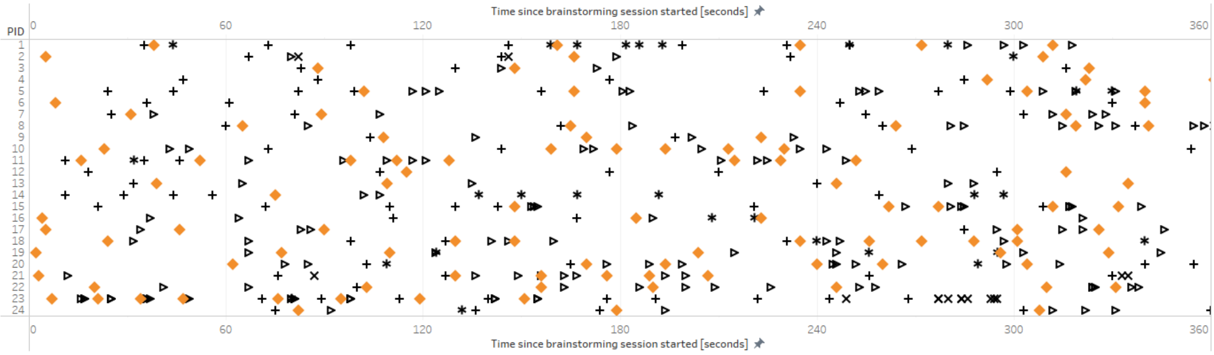

*Figure 3. Timeline for brainstorming sessions (+ added idea, * adjusted node, ♦ requested AI suggestions, ▷ accepted/copied AI suggestion)*

4.2 Cognitive stimulation/synergy

Cognitive stimulation refers to "utterance[s] of members may contain task related stimuli that elicit new ideas from other members" (Pinsonneault et al., 1999), leading to a performance gain. For the reason discussed above participants had to actively request suggestions, which might act as stimuli. While we asked participants to make at least one request for AI suggestions (to allow the investigation of the perception of working with an AI system), all but one participant performed multiple requests (min=1, max=8, mean=4.5). When asked for their reasoning as to why they *requested* AI suggestions, multiple participants stated that they were curious and expected to receive ideas or inspiration, or that they ran out of ideas (e.g., "I ran out of ideas or for the moment didn't think of any that quickly and then got ideas or inspiration from them" (P05)). When asked about the AI suggestions they *received*, participants explained that they did feel inspired through the suggestions. Participants highlighted that the suggestions helped them to explore new areas of ideas:

- The AI "inspired me to think in different directions" and "I really like it. It helps to broaden the horizon and get inspired to think in differs ways/areas than before. After I took the suggestions I was able to come up with a few additional ideas I didn't think of before. And that is just because the AI inspired me to explore this particular area." (P15)
- AI suggestions "hint[ed] me in some direction of thinking." (P14)

- The AI “gave inspiration to think into different directions (e.g.: in the beginning I mainly thought about ideas on how to educate the public, the AI gave me the idea to come up with more direct ideas/things to do)” (P05)

An example provided by one of the participants (P05) was that the AI suggested an idea regarding using solar panels and the participant then “thought about gaining energy from wind”. The responses show that participants expected to receive new ideas and actually felt to have gained new ideas from the AI. We therefore conclude that the AI system might lead to the process gain of cognitive stimulation.

4.3 Free riding

The process loss free riding is the “motivated, intentional withdrawal of efforts”, where group “members might limit their efforts and contributions by relying on others to accomplish the task” (Pinsonneault et al., 1999). Some participants acknowledged having relied on the AI stating that working with the AI is “easier than thinking alone” (P09) or working without it “would have been too much thinking” (P18).

From literature, several underlying factors are known that lead to free riding including (1) perceived dispensability of one’s effort and (2) diffused responsibility (Pinsonneault et al., 1999). With regard to (1) perceived dispensability, several participants stated that they do not feel to be knowledgeable in the topic and therefore welcomed the support by an AI system (P13). They also felt that the brainstorming task required creativity and they find themselves not to be a “creative person” (P16). These are examples where people might feel dispensable and might therefore tend to free ride.

With regard to (2) diffused responsibility, we asked participants “To what degree do you feel responsible for the results, i.e., final list of ideas (0-100%)?” and asked them to justify their response. We report the results on the responsibility scores (figure 4) separately for five participants (group 2), as these participants reported they did not feel to be able to provide an estimation (P12), provided contradictory information (P21), or provided answers more generally related to sustainability (P1, P2, P8). For the remaining participants, most participants (14) reported a higher felt responsibility (mean=64.7%) as compared to the share of ideas they contributed (mean=41.0%), for two participant the values were aligned, and three participants felt a lower responsibility score as compared to their contribution.

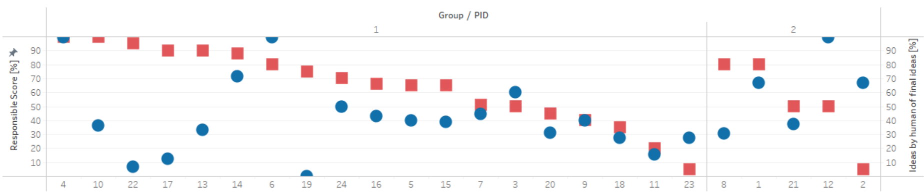

Figure 4. Reported felt responsibility for the final list of results (red squares) vs. share of final ideas by the human (blue circles)

PID	Justification	Explanation	RS	AS
P11	Share of ideas	“I only wrote two or three ideas by myself and accepted a lot more suggestions from AI.”	20%	15.8%
P03	Share of ideas	“The final list consists of half of my own ideas.”	50%	60.0%
P16	Share of top ideas	“My top three choices are two of my own and one AI generated”	66%	42.9%
P07	Curation	“I typed 5 answers from myself and got 5 from AI suggestion system. So I can say 51% because I picked 5 of AI so it was still in my responsability.”	51%	44.4%
P22	Curation	“I chose what to select and what not to.”	95%	6.7%

Table 2. Participants’ justification and explanation for the reported responsibility score (RS), and actual share of contribution towards the final list of results (AS)

Many participant justified their responsibility score with the share of ideas they contributed (see table 2, e.g., P11, P03), while one participant referred to the distribution among the top three choices (P16). Several participants explain to feel high(er) responsibility for the result as they reviewed the suggestions and made the final selection, i.e., curated the results (e.g., P7, P22). Thus, different justifications for the reported responsibility scores exist, with some participants finding it difficult to provide a justification altogether, surfacing the need for further exploratory research perceived responsibility formation.

Consistent with prior findings, the results show participants on average claiming higher responsibility as compared to the share of ideas contributed. Nonetheless, except for two all participants claim lower than full responsibility for the final results, which might hint – again – at a known factor for free riding.

To "free ride" during the task, participants would need to generate suggestions and accept (i.e., copy) them. We thus analyzed the log data: in total 156 suggestions were accepted, only three of those suggestions copied were adjusted by the participants. However, participant's survey responses do not necessarily hint at a withdrawal of efforts, as participants seem to have engaged with the suggestions. Several participants praised the idea quality, as "well thought out suggestions" (P09), which are "perfectly understandable" (P16), "diverse" (P03, P11) and were presented "in the right amount" (P03). They used similar arguments when explaining when and why they accepted suggestions, describing them as "smart" (P17), "good ideas" (P02), "short, precise and I could agree to their message" (P24)". This aligns with their explanations as to why they *accepted* many suggestions *without making changes*:

- "I did not make changes because AI suggestions were good." (P11)
- "Some of them covered blindspots, like composting and rainwater collection" (P20)
- I "did not make changes: because they were clear, and it sometimes didn't make sense to make an idea more specific/less broad" (P05)
- "I did not adjust when they were straightforward okay" (P20)

This high level of perceived quality of suggestions is also reflected at the stage of idea selection. After completing the brainstorming exercise, participants were presented with the final list of their ideas including AI suggestions they had accepted (i.e., copied). Participants were asked to "select a diverse set of your top 3 ideas". While the majority of ideas selected as best originated from a human, 35.2% percent originated from the AI system, with 17 participants (70.8%) selecting at least one accepted AI suggestion as one of their best idea (figure 5). Those final ideas based on AI suggestions included, e.g.,

- "By Introduction of new study programs with a focus on sustainability" (P01)
- "By Establishing a "Green Fund" to support student-led sustainability initiatives" (P21)
- "By implementing waste reduction and recycling programs" (P17)
- "By collecting rain water for reuse in toilet flushing or landscape irrigation" (P20).

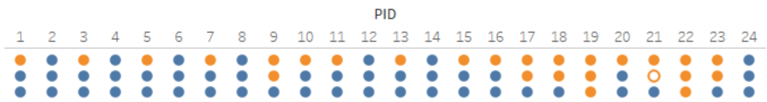

Figure 5. Participants' selected top 3 ideas by origin: ai=orange, human=blue (for P21 one selected idea was based on an AI suggestion after the brainstorming time had passed)

We thus conclude that the high number of accepted, non-adjusted suggestions might at least partially be related to the high perceived quality of AI suggestions and thus might not necessarily indicate free riding. The fact that many of those suggestions were not adjusted by the participants might be partially attributed to the time constraint, and to the fact that, due to the interface design, several participants were not aware adjustments were possible, even though it was explained during the introductions (e.g., P18).

This active engagement of the participants with the suggestions is also reflected in the participants' reasoning as to why the decided not to copy AI suggestions, most notably due to a lack of formal- and content-related quality of the ideas, with grammar errors or ideas perceived to be "basic", "blunt" (P04), or not realistic (P16). With regard to diversity of suggestions, there was disagreement among participants with some praising the diversity (e.g., P03) and others stating a perceived lack of diversity (e.g., P22). On a similar note, there was disagreement with regard to the relative distance of suggestions in relation to the human entered ideas, with some participants stating suggestions were not based on the ideas (e.g., P02) and others stating the suggestions were basically their ideas phrased in different words (e.g., P10).

To summarize, free riding, i.e., the intentional withdrawal of efforts, occurs e.g., when people (1) feel their skills are dispensable, or (2) when responsibility is unclear (Pinsonneault et al., 1999). We see three indications for free riding, with some participants pointing out they did feel to lack qualification for the task (creativity, domain knowledge), most participants' reporting not feeling fully responsible for the results, and participants accepting many of the AI suggestions without making changes. However, the analysis of the qualitative survey responses shows a different picture: participants did critically engage with the suggestions, highlighting both positive and negative aspects around the suggestions and explaining their interaction behavior. According to those responses, the high number of copied suggestions might be due to the good quality of those suggestions. This is particularly underlined by the fact that many of the accepted suggestions were select into their final list of ideas by the participants.

4.4 Observational learning and negative productivity matching

Previous research shows that people within group brainstorming adjust their effort according to the performance of other group members. This can lead to performance gains when members learn from others and try to "imitate best performers" ("observational learning") (Pinsonneault et al., 1999), or lead to performance losses, when group members adjust their "productivity to a baseline level" ("Negative productivity matching") (Pinsonneault et al., 1999). Multiple participants reported to have requested AI suggestions because they wanted to understand what is expected of them by seeing what the AI system would suggest, e.g., to "check if my thinking and ideas were going into the right direction" (P10), to "check if my answers fulfill the requirements of the task" (P15), or "I never thought about this topic and wanted to understand the direction" (P17). We assume that this partly has to be attributed to the study setting, were the participants thought there to be a "correct" way of writing ideas and assumed this would be reflected in the implementation of the AI. However, this might indicate participants calibrating themselves with regard to the AI performance. Thus, providing AI suggestions needs to be done with caution, particularly in the early stages of the brainstorming session if those suggestions implicitly set the baseline for the perceived performance expectations of the human. On the other hand, if humans indeed adjust to the AI, this might open an opportunity to shape humans' performance by careful design.

4.5 Cognitive inertia and cognitive interference

Cognitive inertia refers to the performance reducing effect when individuals "embark on a single train of thought, which limits creativity and productivity" (Pinsonneault et al., 1999). Such an effect occurs during individual brainstorming (nominal groups) as individuals do not receive external stimuli. In our study, participants did not brainstorm with other humans, but they also did not brainstorm alone (but with the AI system). Many participant felt inspiration through exposure to the AI suggestions (cognitive stimulation). However, when asked how they thought the suggestions influenced them some reported that while the suggestions "gave [them] a direction to think about", "they also closed other directions down" (P18) and that the suggestions "made [them] come up with ideas that were close to the ones the AI provided (P05)", with some suggestions being "parallel to [their] ideas" (P07). As mentioned before, some participants felt that the AI re-stated ideas. One participant even explained: "I prefer to first brainstorm alone, and then, only when I'm "stuck" to look at the AI suggestions, because I'd otherwise feel like the ideas of the AI would bias me and prevent myself from being/staying creative" (P05).

On a technical level, the perception of AI suggestions being similar might be explained by the AI system being conditioned by the human input therefore making suggestions related to the input. On this technical level, this could be prevented as part of improved prompt engineering, e.g., by providing a diverse set of sample ideas to condition the AI to "think in different directions". On a more abstract level, the phenomenon of a narrowing of creativity due to stimuli is known as *fixation* (Sio et al., 2015). We thus suggest to employ existing literature on fixation for developing such socio-technical systems.

4.6 Group setting preference: alone, with AI, or with another human

Given their experience, we asked participants if they would have rather worked alone as compared to with the AI, and, if they would rather have worked with another human or with the AI. Most participants preferred working with the AI as compared to alone (79%), stating that "it is nice to have some assistance to be more productive (P07)" and feeling that working with the AI "was easier and faster" (P21) and they were able to "gather[…] many ideas" (P10). However, the AI system does not replace another human, with many participants stating that they would have preferred to work with another human as compared to working with the AI, e.g., due to the "social" aspect (P12) or because they believed it could be more fun (P05, P10). On the other hand, several participants stated they would not have preferred to work with another human, as this would result in "too much discussion" (P18) "tak[ing] up more time" (P11), whereas working with the AI "saves your time" (P01). Several participants stated that they do see working with the AI system and another human as mutually exclusive (P23), but instead suggested a mixture (P03), e.g., the AI systems could be used in a "preliminary step that can lead to a discussion with a human" (P24) or "Having another person to discuss AI suggestions would be nice" (P23).
Participant's preferences, i.e., rather work with an AI than alone, but rather work with another human than with the AI (exclusively), fit the situation explained in the introduction: the AI might improve (perceived) brainstorming performance for individuals but cannot (yet) replace a human teammate.

5 Discussion and future research

5.1 Process effects in human-AI groups

In our study, we propose a human to brainstorm with a generative AI. The setting is similar to all human brainstorming sessions, in that group members contribute as many ideas as possible and build onto other group members' ideas; but different, in that the human requests AI suggests (as this was shown to be most effective, Siangliulue et al., 2015b). We show that certain group effects known to occur within all-human settings (Pinsonneault et al., 1999) also seem to occur in a human-AI setting. Particularly, we saw the potential for the process gain '*cognitive stimulation'*, with participants reporting that they felt they were faster in gathering ideas and covered more aspects as compared to working alone. This is underlined by the large portion of participants wanting to work with the AI instead of working alone. Such a setup could therefore materialize the ambition of increased performance through human-AI collaboration as compared to humans working alone as described by Dellermann et al. (2019). Particularly so, as besides the positive group effect there is a direct effect of the AI system adding ideas. Indeed, participants not only copied many AI suggestions but even selected some of them into their best three ideas, indicating a potential performance improvement. Future research should investigate the qualitative differences between these ideas to improve understanding of human and AI creativity.

However, we also see the risk of the process loss of '*free riding'* to occur. There have been calls for research on free riding in human-AI collaboration scenarios (Siemon and Wank, 2021). Our results are certainly less clear as compared to what Siemon et al. (2015) reported for their study with an AI-like support system, stating no-free riding occurred when using their prototype. A potential explanation could be that the quality of stimuli improved with advances of AI, increasing the participants' perceived dispensability of efforts, a known factor leading to free riding. Another explanation could be the difference in scenario or difference in measurement, with us discussing participants explanations' of their behavior and Siemon et al. (2015) asking participants directly, if they had "exert[ed] less effort" knowing they were supported and their "effort is not instrumental obtained to the outcome". Occurrence of free riding would certainly be consistent with previous research on humans collaborating with virtual assistants (Stieglitz et al., 2022) or on AI assisted decision-making on closed-ended problems, were humans sometimes show signs of "overreliance", accepting AI suggestions even though they are incorrect (Buçinca et al., 2021). This might similarly occur in this brainstorming scenario, particularly given that AI suggestions are not necessarily truthful (Lin et al., 2022). We suggest for future research to investigate how free riding and "withdrawal of efforts" should be conceptualized and measured in these human-AI settings. Humans contributing less ideas when being part of a group is a sign of "withdrawal of efforts" and thereby free riding in traditional all-human real groups. However, in human-AI groups, such a narrow view on the quantity of human ideas might – as our study indicates – draw an incorrect picture, as humans might be engaged in curating AI suggestions instead of developing ideas. A more nuanced explanation might be humans feeling they can reduce their effort to "maintain cognitive resources and enhance efficiency in work" (Stieglitz et al., 2022, p. 758) believing the AI to compensate for them, described by Stieglitz et al. (2022) as "smart loafing". This new role might have contributed to some participants finding it difficult to explain their degree of responsibility for the results.

Overall, we observed signs for both performance enhancing and performance reducing effects. We thus believe it will be important for future research to quantify those performance differences to see the net effect. Typically, brainstorming performance of groups is assessed by measuring quantity and quality of ideas, were quantity is the number of non-redundant ideas (e.g., Michinov, 2012; Ritter and Mostert, 2018) and quality might be assessed, e.g., with regard to usefulness, originality, diversity etc. (e.g., Siangliulue et al., 2015a; Ritter and Mostert, 2018). Measuring performance quantitatively, however, might be unfit here, as the AI can generate an unparalleled number of ideas. Thus, we suggest to adjust performance measurement. As explained earlier with regard to study setup, typically, nominal groups are compared to "real" groups. We, however, argue both from the research path on human-AI collaboration/hybrid intelligence as well as from a practical perspective, ultimately the comparison between (1) human (or another appropriate baseline such as human + search engine or human + ostensible ideas from a social other) and (2) human + AI will be vital. From the research perspective on hybrid intelligence as set out by Dellermann et al. (2019, p. 640) this comparison is important to understand whether the human-AI collaboration actually achieves the goal of "superior results", i.e., if the "socio-technical system achieves a performance in a specific task that none of the involved agents, whether they are human or artificial, could have achieved without the other". Having the human work individually might be a reasonable baseline from a practical perspective, as humans might be faced to either perform the task alone or with the AI, but might not have the option to work with other humans.

Our study offers insights into designing human-AI collaboration for brainstorming and thereby practical implications. We show that generative AI (e.g., GPT-3) may be used to generate helpful suggestions in brainstorming settings. However, such suggestions need to be designed carefully, as they might implicitly be perceived as a performance expectation baseline by the humans. We therefore suggest to make use of prompt engineering to improve suggestion quality both on formal (e.g., grammar) and content (e.g., diversity) level. The latter is particularly important as some participants stated to have found the suggestions to lack diversity, which could lead to a phenomenon called fixation (Sio et al., 2015). Even with those improvements, however, to reduce the human relying on the AI, we suggest to make transparent that the system does not have additional task information. We also suggest careful interaction design to prevent the human uncritically accepting AI suggestions and being invited to free ride. There is much research available on increasing engagement and calibrating trust from AI-assisted decision making (e.g., Buçinca et al., 2021), e.g., through timing or partial disclosure of information.

5.2 Generative AI for collaboration on content-level when solving problems

Generating ideas is an important part of solving open-ended problems. Lots of research was conducted on facilitating brainstorming, e.g., by offering process guidance through technology, more recently with AI enablement. However, whereas meta-level facilitation might be conducted independently of the question, content-level collaboration requires engagement with the specific question at hand. We show that generative AI can enable such collaboration. Due to its flexible design there is no need to anticipate the brainstorming questions or gather training data in advance, as suggestions are generated on the spot.

There are some limitations to be mentioned. We only used one brainstorming question to ensure comparability between participants. While this is a limitation, we want to stress that for our study we did not make any changes to the generative AI (GPT-3) we used. More specifically, we did neither gather training data nor did we train the model. On the contrary, the brainstorming question could simply be replaced in the prompt template string (table 1) to produce different results. We therefore argue that our approach is re-usable and our results might be generalizable to other brainstorming questions, which are sufficiently general and can be answered with public knowledge. That we only used one model (GPT-3) might be considered another limitation. We decided to use GPT-3 as it was shown to have a high performance across many tasks (Brown et al., 2020) and is easily usable via API. Our approach, however, is flexible to be used with other models. Lastly, while we used prompt engineering techniques for prompt construction, results potentially can be improved with more elaborate prompt engineering, potentially even including examples of ideal ideas in brainstorming (Zhu and Luo, 2022).

With our initial study, we offer a transparent, re-usable approach as a foundation. We show that GLMs like GPT-3 have the potential to enable new forms of collaboration for open-ended problems (as opposed to repetitive, closed-ended decision tasks) which are important in organizations (Nickerson and Zenger, 2004), potentially enabling AI to contribute like a an expert/a peer on content-level instead of supporting as a facilitator on meta-level. This might open up the opportunity to support humans in a creative activity, which seemed to be one of the core competencies of humans (Dellermann et al., 2019; Lichtenthaler, 2018). We thereby work towards the research challenge of solving complex problems with human-AI collaboration (Dellermann et al., 2019; Akata et al., 2020; Krogh, 2018). More specifically, Makarius et al. (2020) called for research on "team dynamics of working side-by-side with AI systems", Dellermann et al. (2019) described "superior performance" as a core characteristic of hybrid intelligence, and Seeber et al. (2020) pose the question of how to design machine teammates for effective collaboration. With our study we work towards addressing those challenges in exploring the use of an established lens from all-human brainstorming groups in this new human-AI setting, to better understand team dynamics, particularly performance-affecting processes and ultimately to inform the design of generative AI-based systems to work effectively with humans. While already with this minimal approach most participants preferred working with the system as compared to working alone, we suggest to build on the approach to further improve the collaboration. Besides investigating the robustness of our results, particularly with regard to the limitations (brainstorming question, model, prompt), we propose future research should investigate the integration into settings with multiple humans, existing facilitation approaches on meta-level, and different idea generation techniques.

More broadly, research on this setting, particularly due to its open-ended nature, could inform the discussion around augmentation vs. automation through AI (e.g., Raisch and Krakowski, 2020) and AI-based systems team roles (e.g., Siemon, 2022; Bittner et al., 2019). While cognitive stimulation through the AI resulting in humans developing new ideas might be considered augmentation, the AI adding ideas without human adjustments might be automation. A better understanding of the relative strength of the performance affecting processes (gains and losses) will help to strike a balance to achieve high performance and potentially *superior performance* (Dellermann et al., 2019). Within our setting, human participants took both a contributor (own ideas) and a curator role (selecting from AI ideas). The AI system contributing ideas and thereby working on content-level can be considered a *creator* according to Siemon (2022) team roles. It combines characteristics of both the peer and the expert role according to Bittner et al.'s (2019) team roles. Future research should investigate managing attention (higher attention to others' ideas leads to more ideas generated, Leggett Dugosh and Paulus, 2005) and how this human role change might open up new opportunities for enabling less skilled humans (Rafner et al., 2021), and for reconfiguring (Afiouni and Pinsonneault, 2022) or repurposing jobs (Seeber et al., 2020).

6 Conclusion

With advances of generative, large language models new ways of human-AI collaborations might become feasible. In our exploratory mixed-method study, we investigate humans' perception of working with a generative AI in a brainstorming setting. We show that certain effects know from all-human brainstorming groups seem to appear within human-AI groups too, particularly cognitive stimulation and free riding. We discuss implications for designing such sociotechnical systems and for advancing on the challenge of human-AI collaboration for complex problem solving (Dellermann et al., 2019).

Acknowledgement: This research was funded by the German Federal Ministry of Education and Research (BMBF) in the context of the project HyMeKI (reference number: 01IS20057).

References

Afiouni, R. and A. Pinsonneault (2022). “Ripples of Change – An AI Job Crafting Model for Human-in-Control”. In *Proceedings of the 55th Annual Hawaii International Conference on System Sciences*.

Akata, Z., D. Balliet, M. de Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos, H. Hung, C. Jonker, C. Monz, M. Neerincx, F. Oliehoek, H. Prakken, S. Schlobach, L. van der Gaag, F. van Harmelen, H. van Hoof, B. van Riemsdijk, A. van Wynsberghe, R. Verbrugge, B. Verheij, P. Vossen and M. Welling (2020). “A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence” *Computer* 53 (8), 18–28.

Althuizen, N. and A. Reichel (2016). “The Effects of IT-Enabled Cognitive Stimulation Tools on Creative Problem Solving: A Dual Pathway to Creativity” *Journal of Management Information Systems* 33 (1), 11–44.

Baruah, J. and P. B. Paulus (2016). “The Role of Time and Category Relatedness in Electronic Brainstorming” *Small Group Research* 47 (3), 333–342.

Bittner, E., S. Oeste-Reiß and J. M. Leimeister (2019). “Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work”. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. URL: https://scholarspace.manoa.hawaii.edu/handle/10125/59469.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever and D. Amodei (2020). *Language Models are Few-Shot Learners*.

Buçinca, Z., M. B. Malaya and K. Z. Gajos (2021). “To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making” *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1), 1–21.

Cooper, W. H., R. B. Gallupe, S. Pollard and J. Cadsby (1998). “Some Liberating Effects of Anonymous Electronic Brainstorming” *Small Group Research* 29 (2), 147–178.

Creswell, J. W. (2014). *Research design. Qualitative, quantitative, and mixed methods approaches.* 4. ed., international student ed. Los Angeles: Sage.

Dellermann, D., P. Ebel, M. Söllner and J. M. Leimeister (2019). “Hybrid Intelligence” *Business & Information Systems Engineering* 61 (5), 637–643.

Dennis, A. R., R. K. Minas and A. P. Bhagwatwar (2013). “Sparking Creativity: Improving Electronic Brainstorming with Individual Cognitive Priming” *Journal of Management Information Systems* 29 (4), 195–216.

Dennis, A. R. and J. S. Valacich (1993). “Computer brainstorms: More heads are better than one” *Journal of Applied Psychology* 78 (4), 531–537.

Dennis, A. R. and M. L. Williams (2005). “A Meta-Analysis of Group Side Effects in Electronic Brainstorming: More Heads are Better than One”. In *International Journal of e-Collaboration*.

Di Fede, G., D. Rocchesso, S. P. Dow and S. Andolina (2022). “The Idea Machine: LLM-based Expansion, Rewriting, Combination, and Suggestion of Ideas”. In: *Creativity and Cognition*. New York, NY, USA: ACM, pp. 623–627.

Diehl, M. and W. Stroebe (1987). “Productivity loss in brainstorming groups: Toward the solution of a riddle” *Journal of Personality and Social Psychology* 53 (3), 497–509.

Döring, N. and J. Bortz (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Dugosh, K. L., P. B. Paulus, E. J. Roland and H. C. Yang (2000). “Cognitive stimulation in brainstorming” *Journal of Personality and Social Psychology* 79 (5), 722–735.

Gero, K. I., V. Liu and L. Chilton (2022). “Sparks: Inspiration for Science Writing using Language Models”. In: *Designing Interactive Systems Conference*. Ed. by F. `. Mueller, S. Greuter, R. A. Khot, P. Sweetser, M. Obrist. New York, NY, USA: ACM, pp. 1002–1019.

Hickey, A. M., A. M. Davis and D. Kaiser (2003). “Requirements Elicitation Techniques: Analyzing the Gap between Technology Availability and Technology Use” *Comparative Technology Transfer and Society* 1 (3), 279–302.

Huber, B., S. Shieber and K. Z. Gajos (2019). “Automatically Analyzing Brainstorming Language Behavior with Meeter” *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW), 1–17.

Kohn, N. W. and S. M. Smith (2011). “Collaborative fixation: Effects of others' ideas on brainstorming” *Applied Cognitive Psychology* 25 (3), 359–371.

Krogh, G. von (2018). “Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing” *Academy of Management Discoveries* 4 (4).

Lee, M., P. Liang and Q. Yang (2022). “CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities”. In: *CHI Conference on Human Factors in Computing Systems*. Ed. by S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, K. Yatani. New York, NY, USA: ACM, pp. 1–19.

Leggett Dugosh, K. and P. B. Paulus (2005). “Cognitive and social comparison processes in brainstorming” *Journal of Experimental Social Psychology* 41 (3), 313–320.

Lichtenthaler, U. (2018). “Substitute or Synthesis: The Interplay between Human and Artificial Intelligence” *Research-Technology Management* 61 (5), 12–14.

Lin, S., J. Hilton and O. Evans (2022). “TruthfulQA: Measuring How Models Mimic Human Falsehoods” *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*.

Maaravi, Y., B. Heller, Y. Shoham, S. Mohar and B. Deutsch (2021). “Ideation in the digital age: literature review and integrative model for electronic brainstorming” *Review of Managerial Science* 15 (6), 1431–1464.

Makarius, E. E., D. Mukherjee, J. D. Fox and A. K. Fox (2020). “Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization” *Journal of Business Research* 120, 262–273.

Mayring, P. (2014). *Qualitative Content Analysis - Theoretical Foundation, Basic Procedures and Software Solution.* Klagenfurt.

Memmert, L. and E. Bittner (2022). “Complex Problem Solving through Human-AI Collaboration: Literature Review on Research Contexts”. In: *55th Hawaii International Conference on System Sciences*.

Michinov, N. (2012). “Is Electronic Brainstorming or Brainwriting the Best Way to Improve Creative Performance in Groups? An Overlooked Comparison of Two Idea-Generation Techniques” *Journal of Applied Social Psychology* 42, E222-E243.

Mishra, S., D. Khashabi, C. Baral, Y. Choi and H. Hajishirzi. “Reframing Instructional Prompts to GPTk's Language”. In *Findings of the Association for Computational Linguistics: ACL 2022*, pp. 589–612.

Nickerson, J. A. and T. R. Zenger (2004). “A Knowledge-Based Theory of the Firm—The Problem-Solving Perspective” *Organization Science* 15 (6), 617–632.

Osborn, A. F. (1953). *Applied Imagination. Principles and Procedures of Creative Thinking.* New York: Scribner.

Paulus, P. B., N. W. Kohn, L. E. Arditti and R. M. Korde (2013). “Understanding the Group Size Effect in Electronic Brainstorming” *Small Group Research* 44 (3), 332–352.

Pinsonneault, A., H. Barki, R. B. Gallupe and N. Hoppen (1999). “Electronic Brainstorming: The Illusion of Productivity” *Information Systems Research* 10 (2), 110–133.

Rafner, J., D. Dellermann, A. Hjorth, D. Verasztó, C. Kampf, W. Mackay and J. Sherson (2021). “Deskilling, Upskilling, and Reskilling: a Case for Hybrid Intelligence” *Morals & Machines* 1 (2), 24–39.

Raisch, S. and S. Krakowski (2020). “Artificial Intelligence and Management: The Automation-Augmentation Paradox” *The Academy of Management Review*.

Ritter, S. M. and N. M. Mostert (2018). “How to facilitate a brainstorming session: The effect of idea generation techniques and of group brainstorm after individual brainstorm” *Creative Industries Journal* 11 (3), 263–277.

Salles, A., K. Evers and M. Farisco (2020). “Anthropomorphism in AI” *AJOB neuroscience* 11 (2), 88–95.

Seeber, I., E. Bittner, R. O. Briggs, T. de Vreede, G.-J. de Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe and M. Söllner (2020). “Machines as teammates: A research agenda on AI in team collaboration” *Information & Management* 57 (2), 103174.

Siangliulue, P., K. C. Arnold, K. Z. Gajos and S. P. Dow (2015a). “Toward Collaborative Ideation at Scale”. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. Ed. by D. Cosley, A. Forte, L. Ciolfi, D. McDonald. New York, NY, USA: ACM, pp. 937–945.

Siangliulue, P., J. Chan, K. Z. Gajos and S. P. Dow (2015b). “Providing Timely Examples Improves the Quantity and Quality of Generated Ideas”. In: *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. Ed. by T. Maver, E. Y.-L. Do. New York, NY, USA: ACM, pp. 83–92.

Siemon, D. (2022). “Elaborating Team Roles for Artificial Intelligence-based Teammates in Human-AI Collaboration” *Group Decision and Negotiation* 31 (5), 871–912.

Siemon, D., L. Eckardt and S. Robra-Bissantz (2015). “Tracking Down the Negative Group Creativity Effects with the Help of an Artificial Intelligence-Like Support System”. In: *2015 48th Hawaii International Conference on System Sciences*: IEEE, pp. 236–243.

Siemon, D., R. Li and S. Robra-Bissantz (2020). “Towards a Model of Team Roles in Human-Machine Collaboration”. In *ICIS 2020 Proceedings.*

Siemon, D. and F. Wank (2021). “Collaboration With AI-Based Teammates – Evaluation of the Social Loafing Effect”. In *PACIS 2021 Proceedings*.

Sio, U. N., K. Kotovsky and J. Cagan (2015). “Fixation or inspiration? A meta-analytic review of the role of examples on design processes” *Design Studies* 39, 70–99.

Stieglitz, S., M. Mirbabaie, N. R. J. Möllmann and J. Rzyski (2022). “Collaborating with Virtual Assistants in Organizations: Analyzing Social Loafing Tendencies and Responsibility Attribution” *Information systems frontiers : a journal of research and innovation* 24 (3), 745–770.

Tavanapour, N., D. Theodorakopoulos and E. A. C. Bittner (2020). “A Conversational Agent as Facilitator: Guiding Groups Through Collaboration Processes”. In P. Zaphiris and A. Ioannou (eds.) *Learning and Collaboration Technologies. Human and Technology Ecosystems*, pp. 108–129. Cham: Springer International Publishing.

Venkatesh, V., S. A. Brown and H. Bala (2013). “Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems” *MIS Quarterly* 37 (1), 21–54.

VERBI Software (2019). *MAXQDA 2020.* Berlin.

Wang, K. and J. V. Nickerson (2017). “A literature review on individual creativity support systems” *Computers in Human Behavior* 74, 139–151.

Zhu, Q. and J. Luo (2022). “Generative Pre-Trained Transformer for Design Concept Generation: An Exploration” *Proceedings of the Design Society* 2, 1825–1834.

Zhu, Y., S. M. Ritter and A. Dijksterhuis (2020). “Creativity: Intrapersonal and Interpersonal Selection of Creative Ideas” *The Journal of Creative Behavior* 54 (3), 626–635.