Exploring the impact of AI features on collaborative creativity

JONAS FRICH, Centre for Digital Creativity, Aarhus University, Denmark
JENS EMIL GRØNBÆK, Department of Computer Science, Aarhus University, Denmark
MARCEL BOROWSKI, Department of Computer Science, Aarhus University, Denmark
PETER DALSGAARD, Centre for Digital Creativity, Aarhus University, Denmark

Artificial intelligence (AI) is increasingly integrated into popular digital tools for creative tasks. Many creative professionals also work collaboratively using these tools, and it is therefore important to investigate the effects of AI features on digitally mediated collaborative creativity. Utilizing a controlled experiment with 36 participants grouped in pairs, we assessed the effect of adding AI capabilities to a digital collaboration tool on the generation of creative ideas. The added AI features included support for divergent thinking, sense-making, and convergent thinking. Results, while statistically insignificant, indicate clear tendencies for trade-offs between appropriateness and novelty, such that pairs using AI support developed more novel, but less appropriate ideas. This preliminary study suggests that while AI can enhance the ideation process by expanding the possibility space, it may also lead to a decrease in the suitability of the generated ideas. These findings contribute to the understanding of AI's role in creative collaborations and highlight the nuanced effects of AI tools in creative processes.

ACM Reference Format:
Jonas Frich, Jens Emil Grønbæk, Marcel Borowski, and Peter Dalsgaard. 2024. Exploring the impact of AI features on collaborative creativity. 1, 1 (October 2024), 15 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

1.1 Motivation

In the rapidly evolving landscape of artificial intelligence (AI), one of the most significant disruptions seems to be within the area of creativity. AI technologies are not only meeting, but, in some cases, also surpassing human performance on various creativity assessment tests [5, 9, 12]. Understandably, this shift has sparked a considerable amount of interest and concern regarding the future role of AI in creative domains.

One of the most intriguing challenges is the prospect of augmenting rather than replacing human creativity. This approach suggests a synergy, in which AI-based systems contribute to the creative process without supplanting the human creator. This can potentially materialize in many different interaction modalities—from chatting with AI agents to integrating AI into existing tools to degrees that might not be noticed by the everyday user. A path to adoption that

Authors' Contact Information: Jonas Frich, frich@cc.au.dk, Centre for Digital Creativity, Aarhus University, Aarhus, Denmark; Jens Emil Grønbæk, jensemil@cs.au.dk, Department of Computer Science, Aarhus University, Aarhus, Denmark; Marcel Borowski, marcel.borowsk@cs.au.dk, Department of Computer Science, Aarhus University, Aarhus, Denmark; Peter Dalsgaard, dalsgaard@cavi.au.dk, Centre for Digital Creativity, Aarhus University, Aarhus, Denmark.

Manuscript submitted to ACM

has already occurred is the integration of AI features into widely used digital tools such as MIRO [1] and Figma [2], which are staples in design-oriented creative professions [23].

In many forms of creative work, creativity is often the result of collaborative efforts. However, *we lack knowledge about how systems that integrate AI features influence collaborative creativity*. Addressing this knowledge gap is crucial for developing AI tools to enhance rather than replace human creativity, and for crafting guidelines for effective human-AI collaboration in creative disciplines.

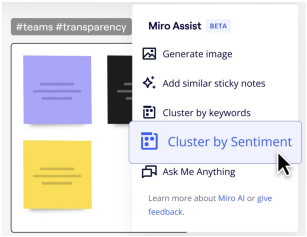

Fig. 1. Miro Assist AI, which includes features for generating and clustering sticky notes, akin to the features studied in this paper

This study constitutes a first step in an investigation of *how human-to-human collaborative creativity is impacted by AI assistance*. We focus on the task of *ideation* for two main reasons: firstly, because it is considered an integral and foundational part of creative processes; secondly, because recent studies have demonstrated that AI-based systems are highly promising for aspects of idea generation [5, 9, 12]. We assess a) the creative process, b) the output, and c) participants' subjective experiences from a collaborative ideation session with versus without AI assistance. The study setup is a pre-registered, between-subject experiment, in which half of the participants had access to three different types of AI features for *divergent thinking*, *convergent thinking*, and *idea assessment*. We examine three central research questions of both confirmatory and exploratory nature.

RQ1: *How are AI features used in the creative process?*

RQ2: *How do added AI features affect the creative outcome?*

RQ3: *What characterizes users' initial experience with AI features?*

1.2 Related work

1.2.1 Creativity. The concept of creativity is vital in human progress, driving innovation and problem-solving across various domains. It fuels advancements in technology, art, science, and culture, making it essential for both individual growth and societal development. While creativity can both unfold in individual and collaborative activities, it generally thrives in environments where ideas can be exchanged and refined through collective effort [19]. Collaboration often entails bringing diverse perspectives to the table, and under the right circumstances, they can converge, enhancing the richness and applicability of creative outcomes.

Measuring and evaluating creativity typically relies on specialized methods or instruments such as the Torrance Tests of Creative Thinking [22] and the Consensual Assessment Technique [1]. These methods depend on explicit definitions of creativity, often involving two key components: *originality* (novelty) and *effectiveness* (value or appropriateness) [17].

[1] www.miro.com
[2] www.figma.com

While these components do not fully capture the richness of human creativity, they provide a useful framework for assessment. Therefore, our study examines the novelty and appropriateness of creative outcomes as the main dependent variables.

We focus on ideation within the creative process, which involves two distinct thinking styles: *divergent* and *convergent thinking* [10, 16]. Divergent thinking is characterized by open-ended exploration and imagination, generating a wide range of ideas, especially in the early stages. Convergent thinking involves selecting, combining, and refining ideas, typically occurring in later stages. Our system targets both thinking styles and includes features for evaluating and selecting ideas, acknowledging the importance of these elements in creative processes.

1.2.2 The role and impact of AI in creative processes. The recent surge in generative AI technology has markedly influenced various creative domains, including writing, video, and image generation with tools like ChatGPT[3], Midjourney[4], and RunwayML[5]. While raising ethical concerns about ownership [3] and inherent biases [18], these recent technological advancements transform the landscape.

In the domain of creative writing, Doshi et al. [7] showed that human writers who relied on AI-generated ideas produced creative stories of higher novelty and interest compared to those written without. Notably, the most creative individuals benefited least from AI assistance, and AI-generated ideas tended to be more similar to each other than ideas generated by humans.

In the more general domain of consulting, a recent large-scale study involving 758 employees from Boston Consulting Group found that AI assistance led to a 40% increase in the quality of results across 18 different tasks that are representative of typical work tasks. A quarter of these tasks were creative tasks such as "(e.g., 'Propose at least 10 ideas for a new shoe targeting an underserved market or sport.')" [6]. These findings suggest that even currently available AI systems can have substantial impacts on both human productivity and creativity. In line with the findings from Doshi et al. [7], the study also found that the introduction of AI acted as a skill-leveler, such that the less capable individuals in the study saw the most significant improvement in performance when they started using AI-based systems.

However, AI might also negatively impact elements of the creative process, in the form of decrease novelty when comparing AI generated ideas to ideas generated by humans [9]. On top of this, idea similarity between individuals is also increased by using AI as demonstrated by Anderson et al. [2], and there is a higher tendency for participants to be fixated on early ideas when using AI instead of existing tools [24].

Given the fact that generative AI chatbots such as ChatGPT and Claude have only been made publicly available in the past two years, there is as of yet limited research on their potential impact on creative work broadly, and ideation specifically. The study from Filippi [8] is arguably closest our study in this paper in terms of the experimental setup. It examines the use and impact of ChatGPT in product design concept generation, with a design brief of creating an innovative pencil sharpener. Half of the study participants were prompted to use ChatGPT, the other half to use conventional concept development methods. The key findings were that the use of ChatGPT led to a higher quantity of concepts, and that these concepts were more novel. However, the designs produced through conventional methods were rated as more useful. Similar findings can be found in the study from Wieland et al. [25], which examined the efficacy of chatbots as brainstorming partners. Participants were tasked with generating ideas for a new product, collaborating either with a chatbot or a human partner. Here, participants who brainstormed with a chatbot generated a significantly higher number of ideas, as well as a greater diversity of ideas. Our study in this paper is similar to these two studies

[3]www.openai.com/chatgpt
[4]www.midjourney.com
[5]www.runwayml.com

in terms of the experimental setups, yet differ from both in that we employ a custom developed prototype in which specific features of ChatGPT are built into the system, and developed such that we can examine their impact in specific phases of the process; in contrast, both [8] and [25] rely on the publicly available standard version of ChatGPT. As AI-based features are increasingly integrated into the systems and services that creative practitioners employ, we find this highly relevant. Moreover, our study specifically focuses on the use of the generative AI features in team ideation, rather than on individuals' use of AI to support ideation.

Other studies have examined co-creative uses of chatbots for ideation, yet with different foci than our study. Hwang and Won [13] examined how a chatbot's perceived identity and conversational style influence team creativity during idea generation tasks. Interestingly, they found that participants produced more ideas of higher quality when they believed they were collaborating with a chatbot, and that the conversational style from the chatbot enhanced creative self-efficacy among participants who typically experienced anxiety in face-to-face group settings. Shin et al. [20] has investigated the potential of chatbots to facilitate consensus-building in asynchronous co-design settings, finding that chatbots were instrumental in increasing participants' willingness to commit to group decisions, even when those decisions ran counter to their personal preferences. Both of these studies are informative for our study, inasmuch as they indicate how conversational interfaces can be designed to support ideation; however, they focus more on participants' perceptions of interacting with AI vis-a-vis human collaborators, and the consequences of this interaction on their creative process. Finally, Tavanapour et al.'s [21] has studied of chatbots for facilitating idea generation in citizen participation, with the intent of articulating and developing principles for designing chatbots that can support ideation. The findings - pertaining to e.g. feedback, conversation style, and task definitions - are relevant for cases in which the chatbot serves as a facilitator, however in our present study the chatbot components are of a different nature, more akin to features and task automations.

Summing up, interaction between human creativity and AI assistance is complex, and as the AI-based systems continue to evolve and become employed in more forms of creative activities, their role in the creative process is likely to become both more significant and nuanced. Our study in this paper adds to our knowledge of how generative AI influences team-based ideation when embedded into collaborative systems.

2 METHOD

In this section, we outline the methodology of our experiment, which is designed to examine the influences of AI-based features on collaborative creativity during the ideation phase. Our study seeks to provide a detailed analysis through a mix of quantitative and qualitative approaches. The protocol for our study has been pre-registered with the Open Science Framework, and a version anonymized for peer review can be accessed[6]. We acknowledge the inherently exploratory nature of this type of research, in which rapid technological advances contribute to a fast changing landscape of systems and tools.

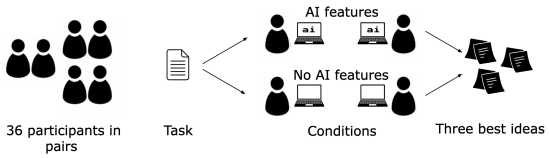

Fig. 2. Overview of experimental procedure

[6] osf.io/uk35d?view_only=9dac3fbcdf844815b8abdd83383bf1f1

2.1 Procedure

We employed a between-subjects design, where groups were randomly assigned to one of two conditions: a control/baseline group (*no-ai* or a group with available an AI features (*ai*). Prior to the experiment, each participant individually completed a short assessment of their verbal creativity and divergent thinking capabilities which also acted like a warm-up exercise. Participants were then introduced to the tasks and the system's basic functionalities (which are detailed below). Following this, participants engaged in a 12-minute collaborative session to solve a creative task, during which the participants were using the system and its features, but were not allowed to access other assistance, e.g. by calling a friend, googling or using other AI tools. At the end, the pair submitted what they deemed to be their three best and most creative ideas.

2.2 Task

We adopted a creative task previously used by Chan et al. [4] on coming up with ideas for a themed wedding, where each idea consists of a theme (e.g. pirate theme) and a prop (e.g. eye patch). Each idea should be a free-form description of the theme and how the prop is used, and by the end of the 12 minutes, participants were instructed to have reached an agreement on the three best and most creative ideas. The participants were also informed that they were free to structure their time as they saw fit.

Fig. 3. Screenshots of the system in the *ai* condition. The panel named "AI Assistant" was hidden in the *no-ai* condition. All other panels were available in both conditions. In this screenshot, the green note is currently selected and its color option and AI annotation is shown in the inspector on the left.

2.3 System Description

To control the collaborative ideation experience, we created a canvas-based collaborative sticky notes application with integrated video feeds for communication. We chose this on the grounds that this form of interface has become commonly used in design, e.g. in the form of Miro and Figma. We provided the same interface in each condition, with the difference being that a dedicated side panel for the AI assistant was enabled in the *ai* condition, and disabled in the *no-ai* condition (see 3).

2.3.1 Sticky Note Application. The core functionality of the system is a sticky note canvas on the right side of the interface. In it, users could create new empty sticky notes by using the "Add Note" button in the top bar or clicking on the canvas while holding the command key. Existing notes could be moved using drag-and-drop while holding the shift key. Clicking on a note selects it and shows an inspector in the left panel that allows to change its color and inspect who created a note. In the same inspector, users could also delete a selected note. Alternatively, they could hold the command key and click on a note to delete it.

2.3.2 Collaboration. The content of the sticky note canvas was real-time synchronized between users and enabled collaboration. Furthermore, the audio and video communication was possible while using the system. A dedicated "Clients" panel left of the canvas showed video feeds of all users using the system,similar to standard video conferencing software.

2.3.3 AI Assistance. In the *ai* condition, an "AI Assistant" panel was added to the "Clients" panel. The AI assistant provided three features to users that could be activated with a single click: (1) An idea generation feature that generated new notes. Notes that were generated by the AI assistant were marked with a small robot icon in the top right corner. (2) A cluster feature that clustered notes into themes and changed the color of notes according to these themes. (3) A rating feature that rated the creativity of notes and changed the opacity of notes accordingly, where less creative notes would be more transparent. This opacity change could be toggled by the user using the "Hide Rating" checkbox. When rating notes, the tool also adds an AI annotation to a note with the explanation for why the creativity was rated in this way. Notes that have such an annotation are marked with a small star icon in the top right corner.

The idea generation and rating features also provided options to customize the tool in the inspector when clicking on the tool. 1 provides an overview of these options.

2.3.4 Implementation. Our system builds on the open-source video-conferencing platform Mirrorverse [11].[7] We adapted Mirrorverse by restructuring the interface to provide more space for the sticky notes and removed features that would not be used in our study in order to simplify the interface. The AI tools were added as a new tool and their interface and options integrated into Mirrorverse.

The AI tools in Mirrorverse connect to a Python server that takes the sticky notes data and options sent from Mirrorverse and formulates a prompt to send to a large language model (LLM). We use the Guidance[8] handlebar templates to create prompts and ensure that the returned values are in a parsable JSON format. Guidance acts as a middleware between our inputs and the LLM. The LLM we used with Guidance is the OpenAI InstructGPT model `text-davinci-003`.[9] We added the Guidance prompts used for all three tools as appendix.

[7] Mirrorverse on GitHub: https://github.com/Webstrates/Mirrorverse
[8] Guidance: https://github.com/guidance-ai/guidance
[9] By the time of writing, OpenAI has shut down and replaced this model: https://platform.openai.com/docs/deprecations/instructgpt-models

Tool	Option	Default	Description
Idea Generation	Count	2	The number of sticky notes that should be generated.
	Context	All Sticky Notes	Existing notes that the tool should include in the prompt. Either all, none, or all notes of a specified color.
	Idea Similarity	Prefer Similar Ideas	Whether the notes that are generated should be similar or different to the ones in the context, or whether the context should not be taken into account.
Rating	Creativity Level	Global	Defines whether the creativity level should be estimated compared to creativity in general ("Global" option) or compared to all other sticky notes on the board ("Personal" option).

Table 1. The options of AI tools available to users in the element inspector panel.

2.4 Participants

38 participants carried out the study in groups of two. Participants were recruited via email, on campus advertisement and by personal reference, and participation was completely voluntary. Participants were either undergraduate, graduate or post-docs at <anonymized>, relatively young (M=23.81 years, SD=2.74), and had average levels of verbal creativity and divergent thinking as measured with the divergent associates test [14] (M=76.90, SD=4.15). Participants were recruited in pairs to ensure that they knew each other and had collaborated before. All participants had taken courses on design ideation, they all had experience in using collaborative canvas systems similar to our prototype, and they also had experience in using ChatGPT.

2.5 Analysis

Quantity and variety of ideas were assessed by counting ideas and categories of ideas. Our analysis strategy for novelty and appropriateness followed that of Chan et al. [4]. Ideas were rated for novelty and appropriateness on a scale of 1-7 by raters recruited on Prolific. We obtained 432 ratings for both novelty and appropriateness for the 54 ideas submitted by the participants. 48 independent raters were recruited on Prolific, each of whom rated a subset of nine ideas including two attention checks to mitigate fatigue. One rater failed an attention check, and an additional rating was done by a new rater. ICC(2,k) scores are 0.89 for novelty and 0.97 for appropriateness, which are on par with existing creativity research as exemplified by the overview of existing datasets provided by Organisciak et al. [15]. DAT scores were calculated using the approach and online services provided by Olson et al. [14].

We qualitatively analyzed the groups' interaction with the AI features using the video and audio recordings of the experiment. Our analysis was interpretive in nature, focusing on understanding the participants' experiences and perspectives. We examined how the groups engaged with the AI features, specifically looking for shifts in their creative process from divergent thinking, which involves generating multiple ideas, to convergent thinking, which involves narrowing down and refining those ideas. We relied on verbalizations from the participants (i.e. "We've got to remember to arrive at only three before the time is up.." Group 1, or "Let's rate them (the ideas)" in Group 2) as indicators of shifts in divergent and convergent thinking. Such shifts were often This qualitative approach allowed us to capture the nuanced ways in which the AI influenced the participants' collaborative ideation processes.

3 RESULTS

Out of the 38 participants, 36 successfully completed the task using the system, of which 8 groups were exposed to the AI condition and chose to use one or more of the AI features. A total of 54 ideas were submitted by the participants, and while some where very novel, but not very appropriate (e.g. "Quentin Tarantino theme. Bride wears yellow jumpsuit, because of the bride in kill bill. Fake blood, Sword from kill bill, Guns, Suitcase with gold"), others were appropriate, but not very novel (e.g. "Flower theme: everyone brings a flower"), fewer were both novel as well as appropriate and thus creative (e.g. "Reverse wedding theme. Everyone wears white clothes except the bride and groom, they will wear black").

3.1 The Process: A potential for more more exploration as AI help to diverge

Overall, we observed that the AI feature to *generate* ideas was by far the most popular (20) compared to the ability to have AI *cluster* (3) or *rate* (9) ideas. We also see that participants with access to the AI tools appeared to have explored more ideas (M=12.88, SD=5.3) than those without (M=9.56, SD=2.07), albeit the difference is not statistically significant using an un-paired t-test t(15) = -1.662, p = .131. This is partly due to a large variation in the AI group, which is potentially explained by the lack of effort required to explore new ideas. This is exemplified by the two participants in group 16, who started their session by using the AI assistant to generate ideas five times within a very short time frame.

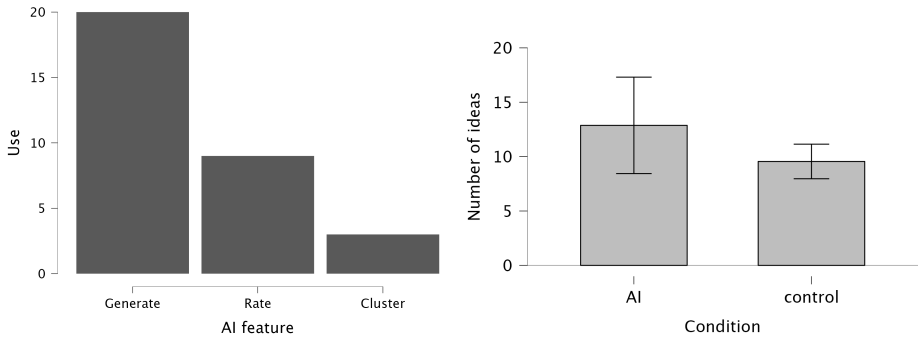

Fig. 4. AI features used and number of ideas across the two conditions

However, participants in both conditions spent the majority of their time on on generating new ideas and content (e.i. divergent thinking) with around 9.5 minutes for both conditions. A qualitative assessment of when and why the participants shifted their focus from one phase to the other in the creative process also reveals that the limited time was a factor for some groups, as exemplified by the P19, in group 10 "I like our ideas, we have like a.. [stops mid-sentence] ..okay we only have one minute left, can we just talk about which ones are our best?". Differences between how groups prioritize their time is clear as P5 and P6, in group 3 around the seven minute mark, with 'only' five ideas in the table discussed the need for moving on to discussing and selecting ideas P5: "What else could be awesome?", P6: "yeah I don't really know, but we only have around five minutes left so we've gotta start choosing three". This dynamic observed in the creative process might indicate that the time pressure added by the experimental setup. In turn, this probably influenced the way in which the added AI features would impact the creative process as offering the potential for more exploration in especially the divergent phase.

3.2 The Outcome: Large but insignificant effects sizes of trading novelty for appropriateness

The analysis of the creative outcome did not reveal a statistically significant difference in the average creativity of the three submitted ideas by the groups in the AI (M=16.05, SD=5.15) as opposed to the control group (M=18.82, SD=5.04), t(15) = 1.121, p = 0.280, Cohen's d=-0.505). This indicates that while the difference in mean test scores did not reach conventional levels of statistical significance, the magnitude of the difference observed was substantial. In other words, we are probably seeing a lack of evidence of an effect of AI assistance on collaborative ideation due to small samples and variations within the data rather than an evidence of no effect.

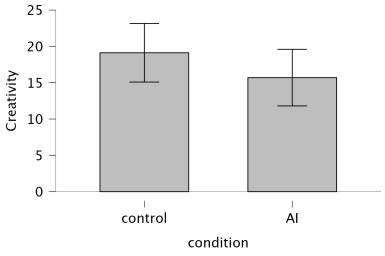

Fig. 5. The outcome of the collaborative ideation on novelty, appropriateness and creativity. Error bars indicate 95% CIs.

The two components of creativity, ie. *novelty* and *appropriateness*, were also not significantly different between the two conditions. Novelty was slightly higher for the groups who had access to an AI assistant (M=4.84,SD=0.54) than those groups that did not (M=4.57, SD=0.46), t(15) = 1.097, p = 0.290, Cohen's d=0.50. On the other hand appropriateness was lower in the groups that had access to an AI assistant (M=3.35, SD=1.07) than those who did not (M=4.11, SD=1.05), t(15) = -1.473, p=0.161, Cohen's d=-0.72). While the differences in the between the two conditions was not statistically significant at the standard interpretation disclaimed in our pre-registration, this first study of AI on collaborative ideation indicates a potential trade-off between novelty and appropriateness in bar plots in figure 6, illustrates that AI potentially has a an effect in a negative direction on the appropriateness (Cohen's d=-0.573) compared to a potential medium effect in the positive direction on novelty (Cohen's d=.0530).

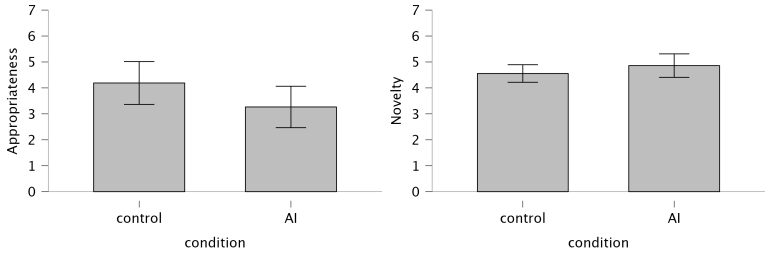

Fig. 6. The outcome of the collaborative ideation on novelty and appropriateness. Error bars indicate 95% CIs.

3.3 The Experience: Insight into participants' experience of using AI

3.3.1 Participants were curious about AI. Across the board, there was a generally curious sentiment towards the what the AI features had to offer. The participants expressed an obvious willingness to employ the features despite being uncertain of what outcome to expect as exemplified by group 8:

P8: I am a bit unsure.. Okay should we try the generate feature? Will you click it or should I?
P7: ..I think just the first.. Generate ideas...
P8: Oh!
P7: [reading an idea aloud] Under the sea... Yes I'll be down!
P8: Okay.. Good ideas

Despite not fully understanding how the AI features worked, the participants demonstrated a readiness to explore its capabilities and showed enthusiasm for the creative suggestions it produced. This indicates a positive initial engagement with AI-driven tools, fostering a potential for increased adoption and integration in their workflows.

3.3.2 AI features as additions rather than automation. Further qualitative analysis revealed that individuals showed a marked preference for the (mostly) generative aspects of AI when participants reached a natural impasse. In several groups, longer periods of silence and lack of new ideas being developed led to participants to suggest the use of the generate feature. (i.e. from Group 14 "Do you want to try this AI tool and see what happens?"). In group 10, the impasse or feeling of being stuck was explicitly stated: "Should I try and press this AI, to see if something else emerges that can get us going again?". This preference may suggest that the creative potential of AI, which allows for the spontaneous generation of novel ideas and content, is valued more highly when a person or group is feeling stuck in a creative process.

3.3.3 Adoption or indirect inspiration. An intriguing aspect of the humans-AI interaction is the manner in which ideas generated by AI are incorporated by the group. Our findings suggest a trend where AI generated ideas are not taken as direct inspiration leading to derived, related ideas. This leads to the question of whether the interaction with AI represents more of an adoption of the AI ideas, rather than inspiration. However, only one out of eight groups included a purely AI generated ideas in as one of their three best ideas submitted by the end of the experiment. This indicates that AI generated ideas were not strictly adopted to replace human ideas in most instances. In the case of Group 2, three existing ideas where on the table, when the *generate* features is evoked and an additional two ideas are added. Because the primary setting of this feature is to include existing ideas as context, the two ideas proposed are thematically similar to those created by the participants, which seems to indicate that they all belong under the same theme "..these are just more for the film theme.. It's a bit basic isn't it?". This statement was followed by a period of silence, after which an idea emerged that clearly broke with the existing, seemingly unconscious theme. It also shows that AI contributions in ideation tasks need not always be clear and obvious, but that they can also take the form of subtle and discreet reinterpretations. This dichotomy raises considerations for when something can be considered inspiration, augmentation or adoption in such a context.

3.3.4 Participants fail to see that the AI mimics their idiosyncratic styles. One notable observation is that participants generally did not pay attention to how ideas generated by the AI mirrored their own unique styling structure. For example participants in group 8 tended to arrange the text in a single sticky note as depicted in the two notes at the top of figure 7, with theme: [name of theme], followed by a new line prop: [name of prop], followed by a new line with a free-form description of the idea, whereas other groups had different styles like bottom two in figure 7 by group 6. The capacity of the AI's to adapt and match the textual style and structure of participants existing ideas might lead to such features being perceived as more intuitive or aligned with the participants own thought processes. However, failing to recognize that the AI matches their personal styling may just be the case of the old saying "when design is good, you don't notice it", highlighting the importance of breakdowns in finding e.g. usability or user experience errors.

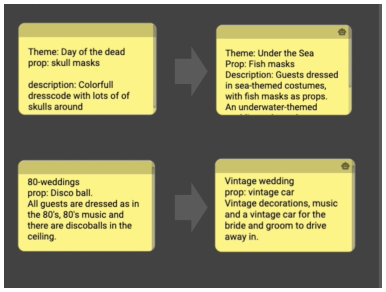

Fig. 7. Caption

4 LIMITATIONS

We consider the main limitation in our study to be the sample size. There are quite clear trends in the results regarding the effects of using vs not using the AI features, with the use of AI features leading participants to generate a higher number of ideas, and these being more novel, although less appropriate. However, these results are not statistically significant, as this would require a larger sample size. The decision to use the given sample size was driven by several factors, mainly centered on our decision to carry it out as an in-person, on-site study. We opted for this in order to control the conditions under which the experiment played out, and to ensure that participants were familiar with working with their collaborator in the study. Recruiting participants for an online study would make it harder to control if and how participants were using other tools than our prototype, and would lower the possibility of realistic collaboration between to familiar individuals. Moreover, previous studies have shown large effect sizes of AI on creativity and productivity, leading us to anticipate similarly significant impacts in our study. Although we did see effects, they were not as pronounced as in these studies. Another limitation concerns the length and complexity of the task at hand. Our task was fairly straightforward and could be solved in a short amount of time. It is plausible that a more complex task would make the effect of AI more pronounced. Taken together, we are encouraged by our results to carry out further studies with more complex tasks for a larger group of participants.

5 CONCLUSION

In conclusion, the integration of AI into popular digital tools for creative tasks is becoming increasingly prevalent, making it crucial to examine the effects of AI features on digitally mediated collaborative creativity. Our study involved a controlled experiment with 36 participants collaborating in pairs. Although the results were statistically insignificant, they revealed clear tendencies for trade-offs between appropriateness and novelty. Specifically, pairs utilizing AI support generated more novel but less appropriate ideas. This preliminary study suggests that while AI can enhance the ideation process by expanding the possibility space and more ideas, it may also lead to a decrease in the suitability of the generated ideas for the problem at hand. We did not match previous effect sizes found in related research, for which reason our findings are statistically insignificant. The pattern of AI assistance leading to more ideas and greater novelty, yet less appropriate for the task at hand, can also be found in the studies from [8], [25], and [7]. Seeing as this is a recurring trend, this prompts further studies to examine how and why AI leads to these outcomes, and how the issue of lower levels of appropriateness with AI support might be mitigated in order to improve meaningful human-AI creative collaboration.

Two additional and pertinent findings emerged from the study. Firstly, participants in the AI conditions clearly preferred the AI assistant to aid them in the divergent activities rather than the convergent activities. While early-stage ideation is important, creative processes often also rely on extensive refinement and development of ideas in subsequent phases, and more studies are required to examine if and how other ways of integrating AI-based features are relevant for these phases of creative work. Secondly, our qualitative analysis suggested that participants exhibited a generally curious and positive sentiment towards AI features and were willing to experiment with the AI tools, even without fully understanding their functioning or potential impacts. This initial curiosity indicates that AI features are seen as beneficial tools that could potentially enhance creativity, although there is still some uncertainty about the best ways to integrate these tools into creative practices. As AI-based tools and features are integrated into evermore digital systems, it will be possible to examine if these positive sentiments persist or are a result of novelty bias.

In sum, our findings contribute to understanding the impact of AI on creative collaborations and underscore the trade-offs that the use of AI-based tools in creative processes currently entail, while also highlighting a number of open questions for the research community to explore.

ACKNOWLEDGMENTS

Parts of this research is funded by the Independent Research Fund Denmark: 0132-00111B, Managing Ideas in Creative Work. We would also like to thank our patient participants for their time and engagement.

REFERENCES

[1] Teresa M Amabile. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology* 43, 5 (1982), 997.

[2] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization Effects of Large Language Models on Human Creative Ideation. *arXiv preprint arXiv:2402.01536* (2024).

[3] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–8.

[4] Joel Chan, Pao Siangliulue, Denisa Qori McDonald, Ruixue Liu, Reza Moradinezhad, Safa Aman, Erin T Solovey, Krzysztof Z Gajos, and Steven P Dow. 2017. Semantically far inspirations considered harmful? accounting for cognitive states in collaborative ideation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. 93–105.

[5] David Cropley. 2023. Is artificial intelligence more creative than humans?: ChatGPT and the divergent association task. *Learning Letters* 2 (2023), 13–13.

[6] Fabrizio Dell'Acqua, Edward McFowland, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R Lakhani. 2023. Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* 24-013 (2023).

[7] Anil R Doshi and Oliver Hauser. 2023. Generative artificial intelligence enhances creativity. *Available at SSRN* (2023).

[8] Stefano Filippi. 2023. Measuring the Impact of ChatGPT on Fostering Concept Generation in Innovative Product Design. *Electronics* 12, 16 (2023), 3535. https://doi.org/10.3390/electronics12163535

[9] Karan Girotra, Lennart Meincke, Christian Terwiesch, and Karl T Ulrich. 2023. Ideas are dimes a dozen: Large language models for idea generation in innovation. *Available at SSRN 4526071* (2023).

[10] Gabriela Goldschmidt. 2016. Linkographic evidence for concurrent divergent and convergent thinking in creative design. *Creativity research journal* 28, 2 (2016), 115–122.

[11] Jens Emil Sloth Grønbæk, Marcel Borowski, Eve Hoggan, Wendy E. Mackay, Michel Beaudouin-Lafon, and Clemens Nylandsted Klokmose. 2023. Mirrorverse: Live Tailoring of Video Conferencing Interfaces. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (<conf-loc>, <city>San Francisco</city>, <state>CA</state>, <country>USA</country>, </conf-loc>) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 14, 14 pages. https://doi.org/10.1145/3586183.3606767

[12] Jennifer Haase and Paul HP Hanel. 2023. Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *arXiv preprint arXiv:2303.12003* (2023).

[13] Angel Hsing-Chi Hwang and Andrea Stevenson Won. 2021. IdeaBot: Investigating Social Facilitation in Human-Machine Team Creativity. In *CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. https://doi.org/10.1145/3411764.3445270

[14] Jay A Olson, Johnny Nahas, Denis Chmoulevitch, Simon J Cropper, and Margaret E Webb. 2021. Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences* 118, 25 (2021), e2022340118.
[15] Peter Organisciak, Selcuk Acar, Denis Dumas, and Kelly Berthiaume. 2023. Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity* 49 (2023), 101356.
[16] Mark A Runco and Selcuk Acar. 2012. Divergent thinking as an indicator of creative potential. *Creativity research journal* 24, 1 (2012), 66–75.
[17] Mark A Runco and Garrett J Jaeger. 2012. The standard definition of creativity. *Creativity research journal* 24, 1 (2012), 92–96.
[18] Teresa Sandoval-Martin and Ester Martínez-Sanzo. 2024. Perpetuation of Gender Bias in Visual Representation of Professions in the Generative AI Tools DALL· E and Bing Image Creator. *Social Sciences* 13, 5 (2024), 250.
[19] R Keith Sawyer and Stacy DeZutter. 2009. Distributed creativity: How collective creations emerge from collaboration. *Psychology of aesthetics, creativity, and the arts* 3, 2 (2009), 81.
[20] Donghoon Shin, Antti Tikkanen, and Antti Oulasvirta. 2022. Chatbots Facilitating Consensus-Building in Asynchronous Co-Design. In *The 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. https://doi.org/10.1145/3526113.3545671
[21] Navid Tavanapour, Mathis Poser, and Eva A. C. Bittner. 2019. Supporting the Idea Generation Process in Citizen Participation - toward an Interactive System with a Conversational Agent as Facilitator. In *Proceedings of the 27th European Conference on Information Systems (ECIS)* (Stockholm, Sweden).
[22] E Paul Torrance. 1966. Torrance tests of creative thinking. *Educational and Psychological Measurement* (1966).
[23] UXtools 2023. *2023 Design Tools survey*. Retrieved May 15, 2024 from https://uxtools.co/survey/2023/
[24] Samangi Wadinambiarachchi, Ryan M Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
[25] Britt Wieland, Jan de Wit, and Alwin de Rooij. 2022. Electronic Brainstorming With a Chatbot Partner: A Good Idea Due to Increased Productivity and Idea Diversity. *Frontiers in Artificial Intelligence* 5 (2022), 880673. https://doi.org/10.3389/frai.2022.880673

A AI TOOL GUIDANCE PROMPTS

Our system uses Guidance 0.0.64 with the initial handlebar templates. The code in the listings was indented for better legibility. The variables used that are shared across the prompts and an example set of the `sticky_notes` variable are listed in the following listing:

Listing 1. Shared variables across prompts. The `sticky_notes` variable depends on the notes that are sent from the Mirrorverse tool.

```
colors = ["Red", "Yellow", "Blue", "Green"]
ratings = ["1", "2", "3", "4", "5", "6", "7", "8", "9", "10"]
task = "Your goal is to come up with ideas for a themed wedding where each ideas consists of a theme

sticky_notes = json.loads("""[
    {
        "uuid": "concept3a08fd36da8b928bdbe7",
        "content": "A medieval theme with crowns and swords.",
        "color": "Red",
        "top": 446.5887756347656,
        "left": 132.28912353515625,
        "width": 240,
        "height": 160
    },
    {
        "uuid": "concept1e0d272e50ef32fd1d58",
        "content": "A western theme with cowboy boots",
```

```
        "color": "Blue",
        "top": 187.7301025390625,
        "left": 136.1640625,
        "width": 240,
        "height": 160
    }
]""")
```

A.1 Idea Generation Tool

Listing 2. The Guidance prompt for the Idea Generation Tool.

```
prompt_idea_generation = guidance("""
There is an ideation process for the following task going on:

"{{task}}"

{{#if context != 'none'}}
    There are the following sticky notes:
    {{#if context == 'all'}}
        [{{#each sticky_notes}}"{{this.content}}",{{/each}}]
    {{else}}
        [{{#each sticky_notes~}}
            {{~#if this.color == context }}
                "{{this.content}}",
            {{/if~}}
        {{/each~}}]
    {{/if~}}
{{/if~}}

Create {{count}} new sticky notes for the given task {{#if idea_type != 'any'}}{{#if context

```json
[{{#geneach 'new_sticky_notes' num_iterations=count join=', '}}"{{gen 'this' stop='"'}}"{{/ge
```
""")
```

A.2 Clustering Tool

Listing 3. The Guidance prompt for the Clustering Tool.

```
prompt_clustering = guidance("""
Group the sticky notes based on their content into themes. Change the color of the sticky notes acco

List of themed notes with changed colors:

```json
[
    {{#each sticky_notes}}
        {
            "content": "{{this.content}}",
            "color": "{{select 'this.classified_color' options=colors}}"
        },
    {{/each~}}
]```
""")
```

A.3 Rating Tool

Listing 4. The Guidance prompt for the Rating Tool.

```
prompt_rating = guidance("""
There is an ideation process for the following task going on:

"{{task}}"

Rate the creativity of the following sticky notes on a scale from 1 (not creative) to 10 (highly cre

```json
[
    {{#each sticky_notes}}
        {
            "content": "{{this.content}}",
            "rating": "{{select 'this.rating' options=ratings}},
            "explanation": "{{gen 'this.explanation' temperature=0.7 max_tokens=40 stop='"'}}"
        },
    {{/each~}}
]```
""")
```