Exploring the Potential for Generative AI-based Conversational Cues for Real-Time Collaborative Ideation

Jude Rayan
jrayan@ucsd.edu
University of California, San Diego
La Jolla, CA, USA

Dhruv Kanetkar
dkanetkar@ucsd.edu
University of California, San Diego
La Jolla, CA, USA

Nicole Gong
y5gong@ucsd.edu
University of California, San Diego
La Jolla, CA, USA

Yuewen Yang
yuewen.yang@nyu.edu
New York University
New York, NY, USA

Srishti Palani
srpalani@ucsd.edu
University of California, San Diego
La Jolla, CA, USA

Haijun Xia
haijunxia@ucsd.edu
University of California, San Diego
La Jolla, CA, USA

Steven P. Dow
spdow@ucsd.edu
University of California, San Diego
La Jolla, CA, USA

Figure 1: (A) The wizard listens in to the active brainstorming discussion between the dyad. (B) The wizard presents ideators with a cue. (C) Participants incorporate the cue into their creative discussion. Image Courtesy: ©icons8.com and ©undraw.io

ABSTRACT

What is the potential value and role for AI to facilitate real-time creative discussions? The paper explores principles for Generative-AI based conversational support by investigating how humans – playing the role of an AI agent – generate contextual conversational cues to guide an ideation session. We studied n=42 people (14 triads) brainstorming through a remote meeting design probe that allows a wizard facilitator to oversee the ideation and send text-based cues that appear real-time in the ideator interface. Thematic analysis of conversations, cues and post-hoc reflections by facilitators uncovered focal points, strategies and challenges. Notably, 44% of the cues sent out by the facilitators were either dismissed or ignored because they did not notice the cue update. When ideators did notice cues, certain facilitator strategies impacted the conversation more than others. Based on our analysis, we present design opportunities to improve generative AI-based systems to better support real-time creative collaborations.

KEYWORDS

Creativity Support, Collaboration, Empirical Methods, Mixed Methods, Prototyping/Implementation

ACM Reference Format:
Jude Rayan, Dhruv Kanetkar, Nicole Gong, Yuewen Yang, Srishti Palani, Haijun Xia, and Steven P. Dow. 2024. Exploring the Potential for Generative AI-based Conversational Cues for Real-Time Collaborative Ideation. In *Creativity and Cognition (C&C '24), June 23–26, 2024, Chicago, IL, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3635636.3656184

C&C '24, June 23–26, 2024, Chicago, IL, USA

ACM ISBN 979-8-4007-0485-7/24/06
https://doi.org/10.1145/3635636.3656184

1 INTRODUCTION

Facilitating verbal brainstorming discussions can be a daunting task. Research on brainstorming shows that ideators often generate the most obvious ideas first, and later struggle to push for truly novel concepts [32, 40, 52]. Facilitators must consistently keep track of the rapidly unfolding conversations and intervene with new angles

or considerations that could trigger novel ideas [15–17]. Given the difficulty of facilitating brainstorming, or even finding skilled facilitators, to what extent could (or should) Gen AI (Generative AI) play a role in supporting verbal ideation sessions? Large language models are now capable of quickly comprehending and generating language on the fly [19, 59, 71], but they typically require significant attention and iteration to glean valuable outputs [75]. To understand the potential role and desired interactions with Gen AI, our research explores how ideators react to different conversational cues.

Researchers have long explored how to supplement meetings with the help of AI, such as including conversational analytics [57], text-based summaries and transcripts [6], relevant visual references[42] and visualization support [5, 64, 66]. For instance, Lyon et al. introduced the idea of 'dual-purpose speech' – systems that listen to verbal conversations and implicitly access related information such as calendar information and read out contextually relevant information from a prior conversation recording [44]. Crosstalk is another example of a voice-based meeting support tool that uses intent recognition to proactively retrieve contextual information from the internet and local files for uninterrupted communication during collaboration [76]. Likewise, Andolina et al.'s system, 'InspirationWall' [4] listened to an ongoing conversation to extract and display words in a floating manner on a shared display to catalyze idea generation. Drawing inspiration from prior work on augmenting live discussions, our research seeks to understand how and what types of conversational discussion cues might be generated by a human facilitator during brainstorming and how this impacts an ongoing conversation. Building on this foundation, our study explores how human facilitators role play as an intelligent agent to gather empirical data that could inform the design of real-time AI facilitators for ideation. Our research questions are:

- (RQ1) What kinds of conversational cues are generated by human facilitators (role-playing as AI) during verbal ideation sessions?
- (RQ2) How do different conversational cues impact the ongoing creative discussions?
- (RQ3) What insights do the human facilitators give for generating machine-generated cues?

To explore these questions, we hosted a virtual meeting and created custom interfaces to support interactions between pairs of ideators (dyads) and hidden facilitators (wizards). The wizard listened to the ideating conversations, typed, and sent out different text-based conversational cues. The ideators can instantly view the cues sent out by the facilitator and use them to guide their conversations or ignore them. The interface also hosted a real-time collaborative note-taking space for ideators to jot down ideas, and thoughts during their discussion. We conducted a between-subjects study (n=42), where we recruited three participants for each session (two ideators and one human wizard facilitator). The ideators were asked to brainstorm potential solutions for the problem of bullying in schools, while facilitators generated text-based cues throughout the conversation. We observed a total of eight different strategies (including combinations) ranging from helping ideators uncover deeper dimensions to reminding them of the task logistics and objectives. We compared each strategy in terms of ideators' perceived helpfulness and timeliness and the probability of having conversational impact. We find that each strategy showed differences in ideators' perceived helpfulness, timeliness and the probability of having a conversational impact. The most used strategy was 'Dig Deeper into Existing Dimension', the most helpful and timely strategy was 'Introduce New Dimension'. The strategy with the highest conversational impact to no impact ratio was also found to be 'Introduce New Dimension'. We also share qualitative insights on how ideators and facilitators reflect on their experiences during the course of our study. Ideator insights include preferences for how and what types of conversational support ensure good collaborative ideation experience. Facilitator insights touch upon their facilitation logic, and challenges encountered during the sessions. We discuss the potential implications for facilitating collaborative ideation, both for human and machine-supported contexts. We offer the following contributions:

- A prototype of a virtual meeting platform that enables brainstorming and provides real-time conversation cues (via hidden Wizard-of-Oz facilitators).
- Empirical insights on the different types of conversational cues that could be generated and their impact on creative production and perceived helpfulness and timeliness.

2 RELATED WORK

This paper builds on prior research on group brainstorming with facilitators, advancements in Generative AI to support creativity, and developments in meeting support technology to explore the potential of real-time AI-based collaborative ideation support.

2.1 Group Brainstorming and the Role of Facilitators

Brainstorming is a form of divergent thinking to generate multiple creative pathways towards reaching an objective while leaning on strategies such as approaching a problem space from a new perspective [39] or coming up with alternative uses [29]. Group brainstorming seeks to harness the diverse perspectives of multiple individuals to enhance the overall creativity of the ideas shared [53]. Sanders et al.'s model for co-creation postulates group brainstorming as a process in which participants share, fuse, and rearrange their idea fragments, thereby giving rise to novel ideas [65]. However, group brainstorming poses a variety of challenges such as coordinating diverse knowledge [26], groupthink [34], and managing the potential influence of various social factors [55]. More nuanced challenges such as evaluation apprehension, production blocking, and free-loading have been found to lead to drops in productivity [21].

The role of facilitators is to strategically enhance collaborative ideation while alleviating the potential social challenges to maintain a meaningful discussion. Experienced facilitators employ a wide variety of strategies to support creativity during collaborative ideation. Encouraging constructive conflict [20] and cognitive stimulation [49] are two popular approaches, both functioning to trigger divergent thinking. Researchers have tried to integrate some of these facilitation strategies in systems to support collaborative discussions. For example, Momentum provides thinking cues to the ideators before the meeting, in order to help them stay on topic

and contribute better during the brainstorming session [7]. Another example within the learning context is a virtual robot agent by Mizrahi that was trained on pre-determined scenarios during learning modules to facilitate discussions between learners [46].

While pre-determined scenarios can provide high-level perspectives, ideal facilitation support would be responsive to the real-time rapidly unfolding conversations and tailored to probe ideators to think about deeper dimensions of their idea. Researchers have compared the utility of different idea generation techniques such as "Silence, Evolution, Random Connections, and Scamper", and found the number of ideas generated to be the same, but observed differences in each technique's flexibility and usefulness [62]. Many systems demand explicit interaction with virtual agents to catalyze discussions such as chatting with a bot to motivate student learning [70] and speaking to agents in pre-planned settings for language practice [48]. However, as Campos et al. pointed out, explicitly interacting with virtual agents still poses challenges such as managing the right amount of conversational history to maintain the flow of conversation [14]. Similarly, a literature review by Zheng et al. identified similar issues brought up by groups chatting with virtual agents in various collaborative contexts including not capturing the right context, abandoning users, not being able to gain control of the facilitation session, and challenges with striking a balance with the amount of support [78]. These issues can disrupt the very conversation it is intended to support. Our work positions the AI facilitator to not strictly rely on explicit interaction, but rather passively listen and dynamically facilitate discussions during ideation. As mentioned earlier, categorizing different cue strategies has been studied before in group ideation facilitation literature and other contexts such as problem-based learning [31] from a theoretical sense. We were interested in looking at the same but from an empirical perspective in an open-ended in-situ setting, to combine the already existing knowledge from theoretical approaches to holistically guide how to support technology-driven group ideation facilitation.

2.2 Generative AI is a Catalyst for Creativity

Generative AI has opened up a plethora of opportunities for seeding creativity. Its potential to supplement user workflows has been explored in multiple domains including story writing [71], journalism [59], video generation [63], music generation [3], and image generation [60]. AI Chains demonstrated that specialized prompting setups can dampen undesirable language model behaviors such as uncontrollability and low quality responses [75]. Other examples of innovative interaction techniques to control LLM behavior include Talebrush which uses a novel technique by using line sketching to dictate the narrative flow to the language model to generate stories that align with the author's vision[19]. Similarly, PopBlends utilizes the concept of conceptual blending to support the sensemaking of blending two cross-domain entities [74]. Graphologue introduced the idea of diagrammatically interacting with language models instead of using textual prompts [35]. Removing the need for interaction, even autonomously operating agents powered by language models have been able to simulate a set of human actions for different social contexts [51]. Furthermore, co-creation [43] and co-design [41] alongside helpful generative AI agents add to the evidence of how it can contribute in a helpful manner to human collaboration settings. When Gen AI is situated amidst a group discussion, there needs to be a careful balance between when and what kind of inspiration it generates. Our study reports on the different strategies that human facilitators would use in such settings along with qualitative insights on ideator preferences and challenges when working with Gen AI in brainstorming context.

2.3 Facilitating Collaborative Ideation Meetings with Technology

Advancements in speech recognition technology have paved the way for building sophisticated meeting support tools. These support tools have made it possible to understand linguistic insights during a brainstorming discussion [33], track meeting attendee participation [8], provide specialized support for note-taking and capturing important information in meetings [2, 18, 36], voice-driven retrieval of contextual information from the web and proactively initiating in-situ online conferencing actions to support seamless collaboration [76]. Other research has explored how visualizations can supplement meetings: enabling users to revisit and extend prior moments in a conversation by showing visual clusters of key points[9], representing the various points discussed during the meeting to reduce the burden on memory [67], and displaying relevant visual stimuli during conversations [38, 42, 76]. Ideation support in meeting systems plays a more specific role in meetings by directly helping users generate high quality and novel ideas during the conversation. Providing relevant pictorial cues during the conversation also has been found to increase creativity [73]. Automatic clustering of similar ideas and suggestions in a map-like structure enables easier understanding of the design space, inherent relationships between proposed ideas thereby increasing diversity and quantity of ideas [68]. Likewise, the Ideation Compass shows topic-based visualizations to know who contributed to the divergent and convergent parts of the ideation discussion [72]. In this paper, we explore the potential for real-time facilitator generated support cues during ideation with an eye towards realizing the new potentials enabled by LLMs.

3 DESIGN PROBE

This section covers the findings from early piloting sessions to observe in-situ brainstorming behavior that users exhibited in traditional video conferencing applications. These findings were instrumental to designing and developing a design probe capable of displaying contextual support cues.

3.1 Observations from Pilot Studies

Our study focuses on the impact of conversational cues in group brainstorming. We recruited pairs of individuals (dyads) to participate in a version of the alternative uses task [28] where they had to choose one contemporary issue involving college students and ideate solutions for it using Zoom [1]. The goal was to understand how dyads brainstorm in a video conference setting and what affordances could support a productive session. The participants were given a brainstorming task and were instructed to select the best

[1] https://zoom.us/

idea by the end of the session. They were allowed to use a Google Doc to track notes and unsurprisingly, every dyad opened a Google document[2] for note-taking in a shared space so that both members of the team were aware of the ideas and could edit at the same time. Secondly, the ideators did not use the chat interface except for sharing links, signaling a preference for a shared note-taking space to jot down ideas over the regular chat interface. We tried using the Alternative Uses task [28] to pilot the study with the objective of selecting the best idea. However, it often led to the premature conclusion of the discussion due to the simplicity of the task. Therefore, we wanted to design a task with a topic that was serious, required more deliberation, and demanded effort put towards divergent and convergent thinking.

3.2 Design Recommendations for Wizard Facilitation

The research team observed multiple facilitated virtual discussions both small and large groups on the Internet. Then the team convened and shared notes to prioritize different features for a design probe that would allow us to explore AI facilitation through a wizarding approach.

- **Visibility of the Ideating Team**: It is important for the facilitators in virtual discussions to access both audio and video streams during the meeting. Additionally, the facilitator must be able to use their full human ability to anticipate the conversation. Therefore, our wizard (human facilitator) needed to have oversight on the discussion.
- **Hidden Facilitation**: Since we want to simulate an AI agent, the facilitator must be hidden from the ideators. Facilitators need to provide conversational cues to the ideators, but using the regular chatbox would reveal them.
- **Minimal Distraction**: The conversational cues must be easily visible to the participants without distracting the conversation between the ideators.
- **Tools to Support the Collaborative Process**: The ideators need to be able to work on a shared document that updates in real-time in order for them to collaboratively take notes and have awareness of the contents of the document. This shared document has to be situated in the meeting platform so no time is wasted with accessing and retrieving the document from another third party service.

3.3 Probe Interfaces

To achieve an understanding of what kinds of support cues human facilitators would send during a virtual brainstorming session, we needed to observe how human facilitators work in a video-conferencing setting. Inspired by prior work leveraging the Wizard-of-0z technique to study cooperation or collaboration in GUI design [47], pedagogical methodology [37], conversational agents [69], live immersive embodied storytelling environments [24], and iterative design workflows [23], we adopted a method for studying group ideation in this setting. We detail the different views of the design probe, the UI for the post-task, and details about the technical implementation.

[2]https://www.google.com/docs/about/

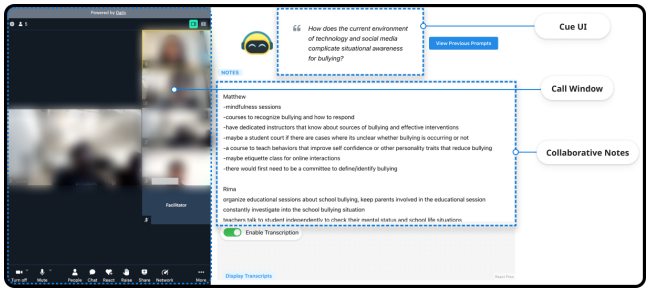

Figure 2: The Ideator's View provides the call window on the left allowing ideators to communicate with each other. On the right, the Cue UI displays the cues sent by the facilitator and the collaborative note-taking space allows ideators to jot down ideas. Image Courtesy:©icons8.com

3.3.1 ***Ideator View:*** We developed a web browser interface that integrates a video conferencing tool, a collaborative note-taking space, and the Cue UI to display the cues generated by the facilitator. The meeting interface displays a call frame on the left side of the screen (See Figure 2) that provides intractability with common features such as voice and video control, chatting, and screen sharing. The right side of the interface houses a real-time collaborative note taking space and the Cue UI component where the cues are displayed. The 'View Past Cues' allowed ideators to revisit prior cues if they had missed any. In order to have a non-intrusive cue notification mechanism, the Cue UI automatically updates once a new cue has been sent. On the bottom right of the screen, there is a button to enable transcription into the database as individual messages attached with metadata.

3.3.2 ***Facilitator View:*** The facilitator view is designed to allow the facilitator to send the cues. Similar to the participant view, the call frame occupies the left half of the interface, allowing the facilitator to listen to the participants' conversations. On the right half of the screen, we provide a text field that allows the facilitator to type in cues (See Figure 3). To ensure that cues sent earlier from the facilitator are not editable after they've been sent, a grayed out text box is created with the submitted cue. The facilitator can view what was sent to the participants by scrolling up within the cue window UI.

3.3.3 ***Post Study View:*** The post study view for the ideators included a scrollable transcript of the conversation with the delivered cues embedded chronologically on the left side of the interface (See Figure 4). On the right side of the screen, the participants can provide ratings on the helpfulness and timeliness of each cue and to describe how they integrated the cue into their conversation. The facilitator used the same setup, but was asked a different set of questions (i) Why was this specific cue sent to the ideating team at this time during the conversation? (ii) Once this cue was sent to the team, did the team use that cue the way you expected they would? If yes, why? If no, why not? The ideators and the facilitator were asked to think aloud while answering the questions to gain deeper insights into their reasoning.

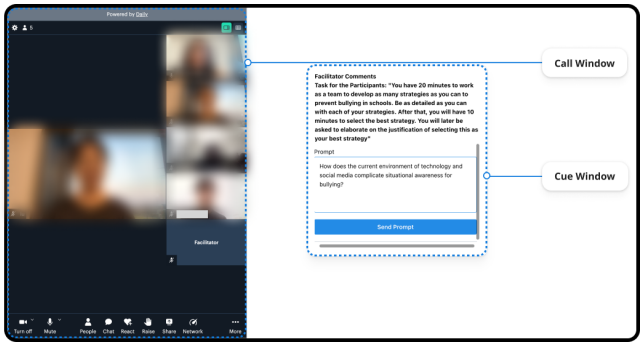

Figure 3: The Facilitator's View has the call frame on the left allowing them to listen in to the ideator conversation. On the right, the Cue Window can be used by the facilitator to type and send new cues to the ideators which would update in real-time for the ideators

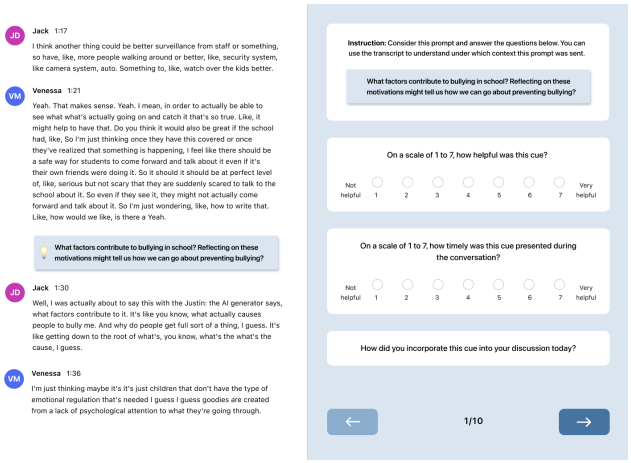

Figure 4: Each participant reviewed, rated and commented on all of the conversational cues provided during the course of the discussion. The number '1/10' signifies that the current cue is the first of ten cues sent by the facilitator

3.4 Implementation

We developed the meeting interface using the React.js[3] on the client side and the mantine component library[4]. We used the Tip-Tap editor[5] in conjunction with y-js[6] to establish a web socket connection, allowing for real time collaborative note taking between participants. To support virtual conferencing, we use the Daily.co API[7] to render the pre-built video call frame. We also use the transcription service from the Daily.co API in order to save the participant conversation for the post-study surveys. To save all conversation transcripts and cues submitted by the facilitator during user studies, we use the Firebase real-time database service [8]. During data collection, we anonymized all user data by generating randomized message and cue IDs. We used Github Copilot [9] during the development process to help with the diagnosis of errors and/or bugs.

4 MIXED METHODS STUDY

We gathered rich mixed methods data on the strategies facilitators used during the brainstorming discussion settings as well as their facilitation logic.

4.1 Participants

We recruited a total of 14 triads (n=42 participants, F=22, M=20) with an average age of 23 years. We assigned two participants to be ideators and one to be the facilitator to oversee their respective dyad groups. The facilitators needed to have at least some prior experience facilitating group discussions or brainstorming sessions. If the triad comprises multiple ideators with prior facilitation experience, then the participant with the most experience will be assigned as the facilitator. Recruitment of participants took place through email, Slack channels, and Twitter for wider reach. We excluded one study session from our data due to the facilitator not fully understanding the task. We also excluded post-reflection insights from one ideator and one facilitator due to missing recording data from the reflection session. All sessions were conducted online and took one hour. Participants, regardless of role, received compensation of 15 USD for their time.

4.2 Data Collection and Analysis

Two coders conducted the thematic analysis for (i) both the Facilitator and Ideator post-reflections (ii) conversational outcomes for the cues in the study. We did two rounds of open-coding, identified overlap, and checked for agreement. All disagreements were discussed and resolved. A coding schema was then developed to continue the coding for the post-reflections and conversational outcomes. An external coder helped with coding a portion of the facilitator post-sessions following the coding schema that the research team had created.

4.2.1 ***Procedure:*** For each session, the research team initiated a meeting link and invited the participating triad (two ideators and one facilitator). The ideators were moved into separate breakout rooms along with one member of the research team. The respective team member conducted a one-minute baseline creative fluency task [28]. In parallel, another research team member onboarded the facilitator to the call where they described the objective of the facilitator's role, task, and responsibilities. The task for the facilitator was to type in conversational support cues in the facilitator window, whenever appropriate to help facilitate the conversation. Facilitators were informed that they were role-playing as an AI and that their cues would appear on the ideators' interface in real time. The facilitator joined the call under the alias 'AI Facilitator' and was asked to never unmute or turn on their video during the call. Once the facilitator onboarding was complete, the ideators were brought out of the breakout rooms and notified about the 'AI Facilitator' integration. Then, a research team member went

[3]https://react.dev/
[4]https://mantine.dev/getting-started/
[5]https://tiptap.dev/
[6]https://yjs.dev/
[7]https://docs.daily.co/
[8]https://firebase.google.com/
[9]https://github.com/features/copilot

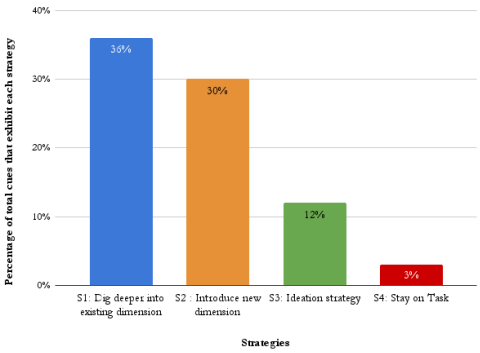

Figure 5: Distribution (in percentage) of cues exhibiting the different strategies that facilitators employed, excluding the combination of strategies. 'S1 - Dig Deeper into existing Dimension' was the most used strategy and 'S4 - Stay on Task and Time' was the least used.

over the key features of the design probe. Then, the following task was introduced to the ideators: **Task cue:** "You have 20 minutes to work as a team to develop as many strategies as you can to prevent bullying in schools. Be as detailed as you can with each of your strategies. After that, you will have 10 minutes to select the best strategy. You will later be asked to elaborate on the justification of selecting this as your best strategy". This task was inspired from [33].

After the meeting, ideators received a link to the Post-Study view (See Figure 4). This survey included the transcript of the meeting with the cues that were sent by the facilitator embedded chronologically with timestamps. Ideators were asked to rate and reflect on the helpfulness, timeliness, and how they integrated each cue sent by the facilitator. Then, each ideator completed a think-aloud Qualtrics survey where they were asked to reflect on the overall brainstorming experience and the AI suggestions. Facilitators also followed a similar sequence of steps. The ideators and facilitators were asked different questions inline with their role in the study. Once the study concluded, the ideators were notified that the discussion cues were generated by a human facilitator not by an AI.

5 FINDINGS

5.1 RQ1: What kinds of conversational cues are generated by human facilitators (role-playing as AI) during a verbal ideation session?

In order to develop an understanding of the different strategies that facilitators employed, we did a thematic analysis of the cues sent during each brainstorming session. A total of 123 cues were collected and analyzed. An average of 8.9 cues were sent by the facilitators. As examples of cues sent by facilitators, F1 posed the question *"It's great that you are thinking about what to want to teach them! How early do you think should kids learn about what bullying is?"* while F3 asked their ideators *"It might help to be specific about some of the strategies, as much as we can. What training strategies exactly can help?"*. According to our thematic analysis, 36% of the cues used the 'Dig Deeper into Existing Dimension' strategy (S1), 30% of the cues used the 'Introduce New Dimension' strategy (S2), 12% of the cues used the 'Ideation Strategy' (S3), 3% of the cues used the 'Stay on Task' strategy (S4). The remaining 19% employed multiple of these four main strategies (See Fig. 5). We also coded 'Tactics' that facilitators used in conjunction with these strategies and found that (T1) 'Framing the Cue as a Question' was the most used with 42.3% of the cues, (T2) 'Referencing Prior Parts of the Conversation' was used in 4.1% of the cues, (T3) 'Providing a Phrase of Encouragement' was used 7% of the time, (T4) 'Referencing an ideator's name' was never used alone but in conjunction with another tactic(s). The remaining 46.6% were combinations of these tactics. 29% of the cues did not use any of these tactics.

The sections that follow show definitions and instances of each cue strategy employed in the study. We situate the examples within its context of use and conversational outcome, and include reflections by the facilitator and ideators.

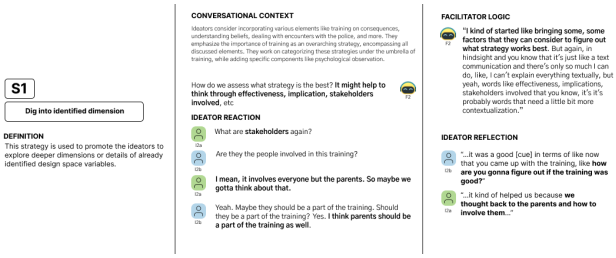

Figure 6: An example of Cue Strategy 1 (Dig into identified dimension). The facilitator used S1 to help ideators increase the specificity of their idea by nudging them towards thinking about evaluation and identifying key participants. Both ideators found the strategy to be helpful in further refining their idea(s). Image Courtesy:©icons8.com

5.1.1 ***S1 - Dig deeper into the existing dimension:*** This strategy aimed to help ideators uncover deeper dimensions of already identified design space variables. In this paper, design space variables are all entities that are a part of the design space that is being considered. For example, in the context of schools, then students, teachers, buildings, artifacts and even abstract entities such as policies would fall under the design space. In this specific example, the context of the conversation includes multiple elements of an idea that revolves around a training component that can be used to counter bullying in schools. Facilitator No.2 (F2) observed that the ideators had discussed how this training component can be delivered, and so F2 probed the ideators to evaluate the "effectiveness, implication, and stakeholders involved" of training strategies. Though the ideators did not use this to guide the evaluation of their ideas, they quickly realized that they overlooked the importance of of a key stakeholder: parents. There were several other instances where facilitators exercised this strategy. For example, F4 sent a cue "Within the school community, what stakeholders can you think of?" as they wanted to "*get them to focus more on the school community specifically because [they felt] lawmakers, law and policy*

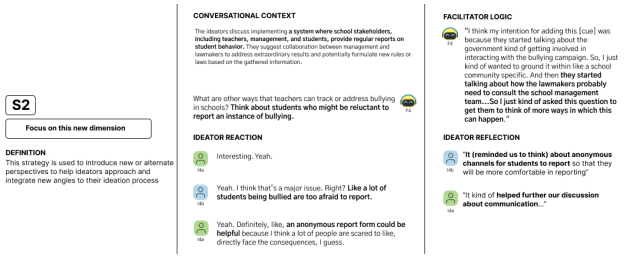

Figure 7: An example of Cue Strategy 2 (Focus on this new dimension). The facilitator used S2 to help the ideators consider a crucial dimension the ideators overlooked in their ideation process. The ideators recognized this was an important dimension and reported actively working on ideas that were cognizant of this dimension. Image Courtesy:©icons8.com

[was] a bit too general". Similarly, F2 sent the cue "What factors contribute to bullying in school? Reflecting on these motivations might tell us how we can go about preventing bullying?" because they did not want the ideators to"*just jump into the solution [but] start talking about motivations of what even causes bullying.*"

5.1.2 ***S2 - Focus on this new dimension/perspective/context:*** This strategy was catered towards helping the ideators focus on new dimensions or perspectives that the ideators can potentially use to approach their ideation process. In the example for S2, the ideators were discussing where information could be gathered between the multiple stakeholders involved in bullying situations. F4 recognized that the trajectory of this conversation was heading towards building information-sharing opportunities between the government, school management, and lawmakers which provoked them to send out the cue "What are other ways that teachers can track or address bullying in schools? Think about students who might be reluctant to report an instance of bullying." This brought awareness to the ideators about potential reporting discomfort making them consider "*anonymous channels for reporting*" Session 4 Ideator B (I4b) and "*further [their] discussion about communication*" (I4a).

5.1.3 ***S3 - Use this strategy to generate more ideas:*** Techniques or frameworks that ideators could use as building blocks for the ideators were another observed strategy. For this specific example, F13 felt that they did not "*have a lot to add on*" and that the pace of the conversation had started to die down. Therefore, in order to "*input more information to see if [it helps] with their conversation*", they sent out the cue "Consider the three R's of bullying: Recognize, Report, and Refuse". An important point to note is that this cue was sent close to the end of the discussion, leaving only a few minutes to react to this cue. This resonated in one of the ideators' comments that it "*offered [them] some new insights, although it did not lead to any direct results*" (I13b). However, the other ideator acknowledged that the cue helped tie the conversation as "*a good conclusion [which also works] for people without any clues, [as] this will be a good starting point*" (I13a).

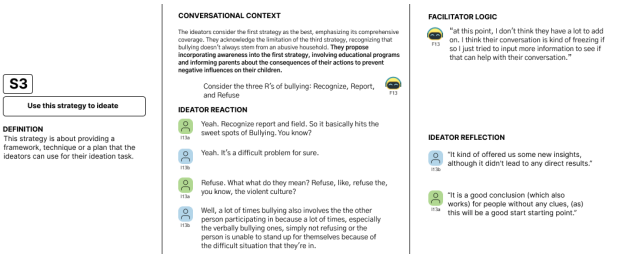

Figure 8: A detailed example where S3 was used. In this example, the ideators were trying to find a concrete angle to ideate towards. The facilitator used S3 to provide the ideators a framework that they could use as a launchpad for their ideation. The ideators had mixed responses about this cue, but also recognized the potential of this cue. Image Courtesy:©icons8.com

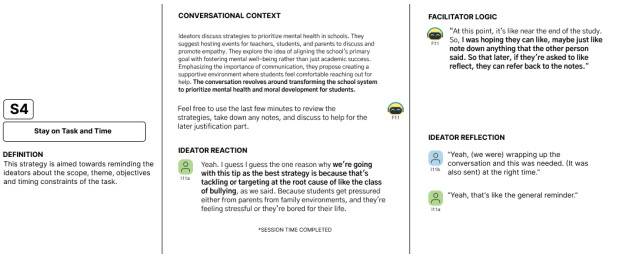

Figure 9: An example where 'S4' strategy was used. In this example, the time to finalize the strategy is about to run out. The facilitator sends a cue using S4 that reminds the ideators that time is about to run out and what they can do in the remaining minutes. The ideators both acknowledge that the cue was sent at the appropriate time. Image Courtesy:©icons8.com

5.1.4 ***S4 - Stay on task and time:*** This strategy is for reminding the ideators about the logistical objectives of the task. In this example, the last few minutes of the conversation revolved around creating more awareness about mental health and moral development for students. But, this had several ideas and not just one. F11 acknowledged that it was near the end of the study and the ideators could leverage notes which could be of aid during reflection time. This provoked F11 to send the cue "Feel free to use the last few minutes to review the strategies, take down any notes, and discuss to help for the later justification" part. Both the ideators understood the intent of the cue that it was a "*general reminder*" (I11a) and was in-line with them "*wrapping up the conversation*" (I11b).

5.2 RQ2: How do conversational cues impact ongoing creative discussions?

This section reports on the different ways in which the different cue strategies and tactics impacted the course of the creative discussion. Based on our analysis of the discussion transcripts, we identified four different conversational outcomes that ensued once the facilitator sent out a support cue: (i) the conversation proceeded

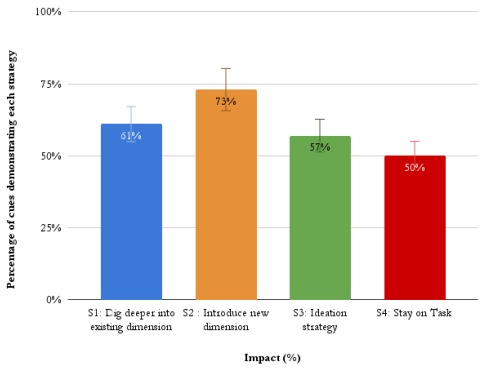

Figure 10: Strategies and their corresponding conversational impact to no impact ratio. The highest impact percentage for individual strategies was observed for S2 (Introduce New Dimension), followed by S1 (Dig Deeper Into Existing Dimension), then S3 (Ideation Strategy) and finally S4 (Stay on Task)

at the same angle as the cue (ii) the conversation went into a different angle for ideation, loosely inspired by the cue (iii) ideators discussed the cue but dismissed it and (iv) ideators did not discuss the cue. We condensed these 4 different outcome possibilities to a binary outcome called 'Impact'. If the ideators discussed the cue and integrated it into their conversation in the same angle the cue was defined in, or they focused on select parts of the cue and created their new angle, these two paths would constitute as 'Impact'. On the other hand, if the ideators discussed the cue and chose to dismiss it or ignored it altogether, they would fall under the criteria of 'No Impact'. We found 56 % of the cues to have had an 'Impact' on the conversation, whereas 44 % of the cues did not have an impact.

5.2.1 ***The strategy of introducing new dimensions was most likely to impact conversation:*** We also computed the conversational impact to no impact ratio for each facilitator strategy (See Fig. 10). The highest impact to no impact ratio was for the Strategy 'Introduce New Dimension' (S2).

In addition to coding the conversational outcomes, we also intended to explore if the strategies embedded in these cues led to different conversational outcomes. Therefore, we performed a chi-squared test to check if there is an association between the type of strategy used by the facilitator and the conversational outcome. The test revealed a significant relationship between the strategies and the outcomes [$\chi = 20.76$; $p < .05$]. Similarly, a chi-squared test was also performed to understand if different tactics play a role in determining the conversational outcome. The results of the Chi squared test denoted a non-significant relationship between the tactics and the conversational outcome [$\chi = 0.655$; $p > .05$].

5.2.2 ***Ideators rated some strategies as more helpful and timely than others:*** At the end of the brainstorming discussion, each ideator from the dyad rated the helpfulness of each cue sent out by the wizard facilitator on a scale of 1 to 7 with 1 meaning strongly disagree and 7 meaning strongly agree. The cues that received a score of 7 had a variety of reasons as to why it got assigned that

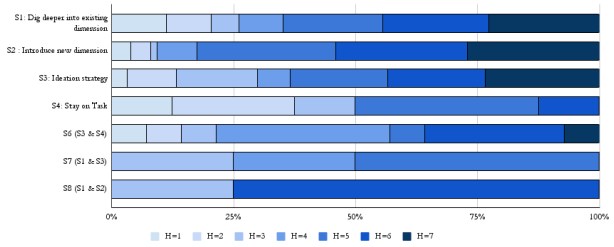

Figure 11: Observed Facilitation Strategies and their corresponding proportions of 'Helpfulness' on a scale of 1 to 7. The most helpful strategies were S1, S2, S3 and S8.

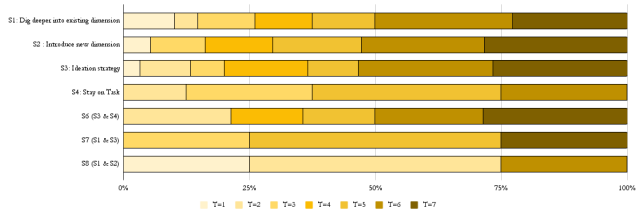

Figure 12: Observed Facilitation Strategies and their corresponding proportions of 'Timeliness' on a scale of 1 to 7. The most timely strategies were S1, S2, S3 and S6.

score. They include the cue introducing an angle that "they were [not talking] about at all" (I2b) all the way to the timeliness of the cue because one ideator "was really about to mention something related to [the cue]" (I2a). On the other hand, cues that received scores in the lower range were due to reasons like the cue being "very common or generic" (I6a), or if the cue "caught [them] off guard" (I7b), or "if [they] were already talking about something else" (I12b). Interestingly, another ideator had missed a cue but while revisiting it, they gave it a lower score because "the cue gave clear guidance and specific guidance but the timing was just off" (I9b). These examples speak to how timeliness also plays a role in the perception of cue helpfulness. The proportion of helpfulness scores varied across strategies (See Figure 11).

5.2.3 ***Ideators shared insights on what impacted their thinking:***

Some cues helped ideators bring attention to other stakeholders: Ideators reported that the cues provided fresh perspectives during the ideation process and that they helped "*reshape and rethink [their] original ideas*" (I4a). One such example is introducing an important stakeholder belonging to the design space that the ideators haven't talked about (I9a). Such cues helped ideators look at the same problem from different angles to prevent them from being stuck. One ideator talked about how brainstorming could lead down a rabbit hole and that the cues were "*like a stop sign and it was good to stop, reflect and go down another path*" (I7b). In the same line of thought, these cues also pulled them back from focusing too much on one particular stakeholder, and make them consider what can other stakeholders do here. A good example is when the cue pushed

them towards thinking about "*authority figures and what they can enforce" apart from only thinking about what students can do*" (I7a). It also introduced new contexts of consideration such as 'cyber bullying', which was a little further from the problem scenario, but the participants revealed that it was a "*really good perspective because [they] just assumed that this happened in school*" (I8a). From an abstraction standpoint, the cues also helped the ideators think deeper about the flexibility of their idea by encouraging them to think about how it can apply to different subgroups of the stakeholders they initially talked about, which the ideators found to be interesting (I3b). Ideators also appreciated the idea of being able to select cues from multiple models at the same time which could allow them to choose from "*creative, conservative, and cues from different perspectives*" (I3a)

Some cues helped ideators uncover deeper dimensions of the design space: The cues not only enabled ideators to think about the "*stakeholders of [the] problem, but also helped [them] think whether [stakeholder] should be involved*" (I9a). Beyond identification of the stakeholders, it provides the basis for the ideators to also think about how these identified stakeholders would communicate with each other (I3b). In a similar vein, ideators also abstracted a stakeholder group down to thinking about the feasibility of the idea for the "*different age groups*" within the stakeholder group (I5a). It also supported an important aspect of problem-solving by helping them think "*about the root cause*" (I7b). Uncovering nuances of the design space also led to new ideas. For example, the ideators considered the possibility of 'students reluctant to report due to fear', which led to the ideators coming up with the 'anonymous reporting/suggestion box' idea (I8b). Once the ideas have been discussed already, the cues helped resurface assumptions that need to be verified, as that forms the basis for certain ideas (I8a). Another outcome that was observed was that it led the ideators to address specific and important questions like "*What is the potential harm?*" (I3a), "*Why would that matter?*" (I9b), "*What are the implications of that?*" (I4b).

Some cues helped ideators increase the specificity of their ideas: The cues helped the participants reset by probing them to "specifically address the problem rather than just getting stuck into the details" (I5a). Additionally, it also helped them "*elaborate on [their] ideas*" (I4b). Another example of the cues helping increase specificity is when it "*forced [them] to think about a specific people group*" for one of their ideas (I7a). One participant suggested that it would be extremely helpful to "*provide a hypothetical situation as it creates a personal aspect*" which pushes them away from thinking about the problem scenario in a broad manner (I7a). In general, the cues allowed participants to communicate their diffuse thoughts better by "*helping word [them] well*" (I7a).

Some cues helped ideators critique their ideas and guide ideation: Beyond providing inspiration, cues helped ideators understand how to better evaluate their ideas because it is clear that they have to "*identify good strategies, but [they didn't] know what makes [something] a good strategy*" (I8a). During reflection, the cues helped participants "*critique over [their] thoughts*" (I9a) and helped them understand that the direction that they were pursuing wasn't the best by nudging them towards "*the more correct and ideal direction*" (I4a). They felt that the cues in general did a good job in helping them stay on track and provided guidance (I9b, I7b). One ideator admitted that they would run into thinking blocks and struggle to think about the next steps but when "*the AI came in, it was really helpful*" (I10b). Towards the closing sections of the ideation session, the cues helped ideators "*understand the flaws of [their] first few ideas and helped guide [them] towards [their] final decision*" (I4a). The usefulness of these prompts as an aid for selecting was also brought up by another ideator who really appreciated these prompts towards the end of the ideation session"*[they] were trying to concatenate all [their] previous idea generation into one cohesive plan*" (I12a). Since there are objectives in terms of the deliverable for the task itself, participants found it helpful for the cues to "*remind [them] about the timing*" (I8a). From an interaction standpoint, it might take effort to craft prompt generative models such as ChatGPT to serve complex needs. However, what piqued the ideators' interest in our system was that the "*cues [it] provided [was] very aligned with what the [ideators] were talking about*" (I12b).

The timeliness of a cue is crucial: When the cues are not sent at the right time, then it will be viewed as a distraction [58]. Examples from their reflections on certain cues reveal a deeper insight into why the timeliness of each cue is extremely crucial for a cue to be considered 'helpful'. For example, P24 rated one of their cues a '7' because they received it "*when [they] were all kind of silent*". Thus hinting at certain situational preferences to see these cues mid-conversation. One ideator (I14b) rated one of the cues a '6' and claimed that it was sent at an appropriate time because "*[the AI] kind of understood the direction of [their] conversation and was able to suggest things before they had considered it*". Cues that received low timeliness ratings revealed recurring concerns with timing. Cues often arrived when the ideators"*already talked about the [cue] before they even popped up*" (I6a), "*[they're] already way past the discussion on one topic*" (I12b) or *"if there was a little bit of delay*" (I13b). An interesting thought shared by I13b was that they would rather see a cue appear early "*than occurring afterward*". In a similar vein, some participants talked about how cues related to the ideation framework strategies, such as the "*three Rs principles should be sent earlier so that [the ideators] can have a better frame to build on*" for early stages of these conversations (I13a). Though cues might be informational by themselves, the utility of the cue is intertwined with not just the content but also the time it was sent. The proportion of timeliness scores varied across strategies (See Figure 12).

Insufficient prior knowledge for the ideator inhibits the ideation experience: Being equipped with enough background knowledge can go a long way in maintaining the health and productivity of the conversation. Lacking the depth of knowledge to carry the brainstorming conversation is an important theme we realized from ideator accounts. Ideators expressed struggles to empathize and understand the "*gap between their environment and the students' environments because [they're] lacking many experiences making it hard to imagine the reality*" (I11b). In order to formulate feasible plans to execute the various ideas discussed during the brainstorming conversation, it is imperative to be able to critically analyze the various components of a solution. However, ideators reported it being "*kind of difficult to figure out what were the best parts of each*

solution [they] had discussed along with seeing which elements of ideas work synergistically altogether" (I12a).

Starting the ideation process was challenging at times: Starting the discussion in itself was an effortful task due to this (I3b, I5a). One participant suggested that it would have been helpful if they "*had a cue at the very beginning of the ideation stage, like a question*" to jump-start the discussion (I6b). It was "*especially harder in the beginning to find an aspect to dive into the discussion*" (I11b).

Some ideators did not have sufficient familiarity with the topic: Some participants expressed not having enough familiarity with the topic of the discussion (I9a, I3a, I8b). Once the ideation process had begun, the next potential challenge that participants had to navigate was to "*[find] the right wording or knowing how to explain what's on [their] mind (I9b)*". Not being able to leverage external knowledge repositories made the ideators feel "*limited in [their] ideation process*"(I8b). On the other hand, they were open to being "*encouraged to draw on things [they've] experienced or read about*" (I5b).

Challenges with using the design probe: Reports of inconveniences around using the design probe included working in noisy environments where one of the ideators "*could not hear each other very well the entire time*" (I6a). The design probe is setup in such a way that the ideators can both take notes and be able to look at the cues that the facilitator sends out. However, some ideators were "*focused on the note taking and the conversation*" that they "*forgot that the [AI Facilitator] was there for a little bit*" (I14b). Similarly, it was not clear to the ideators whenever there was a change in the cue because "*there's like a lot going on on the screen*" (I9b). One ideator admitted that it took some effort to "*pay attention to the [Cue Window] to see if something's updated*" (I12b). Ideators also revealed that they had to sometimes "*[wait] for a new cue" and that it would be nice to* "*modify the cues or ask [the facilitator] directly*" (I3b). An example of a solution for that could be "*implementing a feature where you can click on a (UI element) to generate a [cue]*" (I10a).

5.3 RQ3: What insights do the human facilitators give for generating machine-generated cues?

This section reports on the insights gained through the facilitator post reflection session. An analysis of the facilitators' reflections uncovered the different conversational signals they paid attention to and the challenges they encountered during their facilitation process in the study. It is necessary to look into the facilitators' reflections as they will serve to inform the design of platforms that generate conversational cues automatically.

5.3.1 ***Pay attention to conversational signals:*** It is crucial to understand the various nuances picked on by the facilitator to decide when a conversational intervention needs to happen. Therefore, we looked into what the facilitators had to say about the conversational signals they had looked out for in the post-interview.

Lulls or pauses: (8/14) Facilitators reported that they looked for lulls or pauses to decide when to intervene. They sent cues when they felt that the conversation was "*getting really quiet*" (F4) or that there was a "*lull in the conversation*" (F8). Beyond looking for lulls and pauses as a group, one facilitator reported that even if one of the ideators were talking, and the other ideator "*did not give as long of a response or if it seemed like they were running out of things to say*" (F4), they sent out a cue.

Conversation specificity: (5/14) Facilitators reported that they had the urge to intervene if they felt that the ideators were not being specific enough, or felt the ideas that were being exchanged lacked depth. Facilitators sometimes found that the ideators were just "*throwing out different ideas for how to address bullying. So the [cues] kind of helped them narrow that down*" (F3). More importantly, they found themselves asking questions like "*How long are they spending time on one idea and how specific are they getting?*" (F1). When the conversational trajectory started to get tangential to the objectives of the discussion, facilitators also tried to play sort of a managerial role to "*make sure they finished [with an] answer*" (F9). Overall, the facilitators believe that "*if [the ideators were] being a little vague, [the cue] brings a little bit of specificity*" (F2).

Pace of idea generation: (5/14) Facilitators revealed that the pace of the idea generation process was a strong indicator that helped them decide if the discussion could benefit from a conversational cue at that point. If the facilitators found the ideators to be "*slowing down*" (F4) or "*circling around the same points that have been discussed so far, [they] thought it would be good to intervene and give [cues]*" (F7). Similarly, if the ideators "*kept talking about the same idea*" (F3) or "*stayed on the same subject for a long time*" (F8), it felt like a good opportunity for the facilitators to send a cue. One facilitator even looked for a quantitative signal like the "*number of ideas they're discussing over time*" (F1).

A few other interesting observations: Facilitators also felt the need to send out a cue when the ideators were not strictly working towards the expectations of the task. For example, if the objective of the task was to develop strategies, but "*[the ideators] are not really mentioning strategies but just like having conservation, [that provoked them] to send something to remind them of that*" (F11). Sometimes the facilitators played a corrective role by proactively trying to prevent ideators from making false assumptions, and so they would "*call [the ideators] on that assumption*" (F12).

5.3.2 ***Address challenges with timing, domain knowledge, and phrasing:***

Formulating timely cues is demanding: (5/14) Facilitators expressed concerns over having to quickly come up with a conversational cue that is helpful with the intention to add value to the conversation. They found it challenging to follow the discussion closely and "*try not to repeat what has been said already*" (F7). Sometimes when the facilitator is about to nudge the ideators towards a particular angle, "*they kind of already discussed it or just started discussing it*" (F4). Since our design probe only allows users to type in the text [cues], facilitators who "*don't type that fast*" found it a bit difficult to "*type all that [content]*" (F5).

Topic-Expertise affects the ability to facilitate meaningfully: (7/14) Facilitators reported not having adequate topic-level expertise to facilitate the conversation. They found the early stages of the conversation easier to facilitate, but when the conversation started gaining depth, they had to resort to using Google to seek "*help with*

creating the questions [since they] needed to have some outside help with that" (F10). Though they felt confident about their facilitation skill, they acknowledged that providing the "*right [cues] to get [the ideators] to think more or broadly was kind of difficult*" (F1). Though facilitators had a strong background in the overall topic of bullying, once the ideators decided to integrate another field such as 'policy', "That was something that the [facilitator] did not have a background in" (F3).

Phrasing the cues is challenging: (4/14) Facilitators found that taking up not only the role of AI, but also the persona posed a set of challenges. Ideators speculated if they "*could have given faster [cues], because they wouldn't have cared about the typos*" (F1). The facilitators were "*very conscious of [their] spelling*" and "*had to make sure that [they] did not do any grammatical [errors] to reveal [their] anonymity*" (F7). This is a challenge, especially for non-native English speakers. Some facilitators were also aware of the possibility that spelling mistakes "*might throw the participants off making them know that [it was a wizard]*" (F5).

6 DISCUSSION

Using a Wizard-of-Oz approach [23], we analyzed the different types of cues sent by human facilitators (role-playing as AI) to aid in the process of brainstorming. We identified four different strategies, namely: digging into an identified dimension (S1), calling out a new dimension (S2), suggesting an ideation strategy (S3), and finally (S4) reminding about the scope and logistical details of the task. We also investigated how different cue strategies affect the probability of a conversational impact. We found 56% of cues impacted the conversational outcomes while 44% had no impact. In terms of the perceived helpfulness and timeliness of cues, ideators gave the task-oriented cues (S4) significantly lower ratings compared to the other strategies. On the other hand, S2 was perceived to be the most helpful and timely strategy. Additionally, it also had the highest conversational impact. Furthermore, facilitators provided insight on key conversational signals which included lulls or pauses in the conversation, the specificity of the ideas being discussed, and the pace of the conversation. They also revealed the challenges they faced, including difficulty in sending cues in a timely manner, lack of specific domain knowledge, and phrasing the cues.

6.1 Design Insights for Creativity Support Tools

6.1.1 ***Generate diverse cues by leveraging domain knowledge and creativity insights:*** We aim to explore the potential for AI to support creative conversations in real-time. Support can even begin before the ideation session by suggesting ideation etiquette and strategies for a productive session [7, 30]. During the ideation process, ideators wanted to be able to lookup external information to fill in potential knowledge gaps (I6b, I8b). While LLMs are known to sometimes generate inaccurate responses [77], grounding language model responses with external knowledge bases could help alleviate that problem [56]. Leveraging external knowledge bases to index knowledge could be effective, but the cues need not always have to comprise external information. Several ideators in our study mentioned that they would be open to choosing between different types of cues, such as creative, conservative, and cues that draw upon the ideators' prior experiences and are not necessarily tied to a domain-specific knowledge base. Building on prior work that explored the beneficial impact of involving multiple domains of knowledge in the ideation process, future work can tap into different strategies or sources of wisdom, including how to support creativity in general independent of a domain-specific knowledge base. [10].

6.1.2 ***Improve the visibility and timing of discussion cues:*** In this study, 44% of the cues that were sent out by the facilitators were either dismissed, ignored, or missed. While users appreciate the utility and potential value of cues, quite a few participants reported that they missed the facilitators' cues as they were engaged in discussion and did not notice when new cues arrived. As one ideator mentioned, it took significant effort to constantly monitor the cue window within the UI. Future work could explore different approaches to notifying the ideators of ideation cues without disrupting the current flow of the conversation [50]. For example, maybe cues could be generated and held until an appropriate lull in conversation occurs. On the other hand, missing the facilitator cues could have also been a result of design flaws within the design probe including the cue update notification being not attention-grabbing enough, or the sizing of the Cue UI with respect to the other UI elements. Alternative modalities— such as verbalizing cues or creating image-based cues — could be another angle worth exploring in future work to deliver the cues and make them more visible or impactful during a discussion. Additionally, adaptive and customizable agent feedback mechanisms [13, 27] seem promising as a route for delivering better cue presentation. Our study only comprised of textual cues, but future work could look into the benefits of using multimodal cues during ideation.

6.1.3 ***Support requesting inspiration anytime using different input mechanisms:*** Some ideators expressed that they had to wait for cues from the facilitator and would prefer to ask for certain types of cues at any given time to have additional control over the ideation process. Along a similar line of thought, researchers have discussed the idea of 'proactive-initiated' interactions that take advantage of spatial sense-making capabilities of Generative-AI [45] for voice assistants to decide when to interrupt. Another ideator mentioned wanting to be able to "*tune up the frequency of the prompts*" (I13b). All the ideators made use of the collaborative notes interface due to the nature of the task. Note-taking can also be supplemented with AI to provide in-context recommendations to support divergent thinking during ideation sessions. In fact, cueing directly in the notebook, reorganization of ideas, quick summaries, automatic highlighting, etc. were suggested as potentially helpful add-ons (I3a, I4b).

6.1.4 ***Create generative dashboards for human facilitators:*** Facilitator interfaces could be equipped with AI support that formulates cues which they could then choose to send. This could be especially effective while facilitating large-scale brainstorming sessions and multiple breakout rooms which demands a lot of attention and having to keep track of multiple threads of conversation. Helping facilitators choose and prioritize what cues to send and

identifying gaps to focus their attention on to decide what types of cues to send out could be vital to the health of the discussions. Also, the cue need not be limited to the text format. Creating affordances to allow the facilitator to drop in other forms of inspiration such as relevant data visualizations, images, and video snippets could help add momentum to the conversation. Being able to implement conversational repair strategies to correct errors is also important for the ideators to pay attention to the facilitator [22].

6.1.5 ***Show detailed conversational insights to help improve facilitator performance:*** There are beneficial opportunities for facilitators to hone their skills after brainstorming sessions as well. For example, automated discourse analysis of the brainstorming conversation could yield insights into how effective were some of the strategies that they used and accordingly curate improvements. At the same time, information about their timing could also be a valuable metric for the facilitators to be considerate about. Auto-generated insights into knowledge about the ideators themselves such as their skills, preferences, roles, and responsibilities could be valuable information for future rounds of facilitation. Post-Session analysis that presents identified knowledge gaps, or suggested material for facilitators to look into for future sessions could be very valuable.

6.2 Limitations and Future Work

Our study comprised of 14 triads with 28 ideators and 14 facilitators from diverse design and facilitation backgrounds. It was a short ideation discussion that included people relatively unfamiliar with the topic that the task was based upon. We unearthed many interesting insights from the ideators as well as the facilitators. Future work should explore other ideation tasks, other types of discussions beyond ideation, and other social configurations (e.g., different mixes of expertise, different numbers of people, etc.) to get a full understanding of how these insights generalize to other situations. The reason being, each meeting has a different set of objectives and accordingly needs the right personnel and meeting structure [1]. Though we reported in-depth examples of the strategies to help compare them, we did not delve into further details to do a detailed comparison of each cue strategy. Future work can take advantage of linguistic study methods such as discourse analysis to compare cue strategies and their conversational impact in detail. Furthermore, the personality of the ideators also has the potential to change the dynamic and performance of the meeting [12]. In order to successfully participate in this study, the ideal ideator/facilitator would have sufficient domain expertise/familiarity with the topic of bullying. However, instead of recruiting domain experts as facilitators, we focused on recruiting facilitators who had expertise in leading discussions in design. It is a limitation of the study that we excluded novice facilitators. However, prior research has compared facilitated crowd ideation between experts and novices and found novice facilitators to dampen the creativity of the crowd [16]. Moreover, we were interested in how participants facilitate in a naturalistic setting apart from specifically looking into how the topic expertise of facilitators mediate their facilitation style. Future work in this topic should delve into enquiring the difference in group conversation quality facilitated by experts when compared to non-experts. Though we screened participants to ensure that they were good facilitators, not all of them had sufficient domain knowledge on the topic of anti-bullying strategies. Expertise in facilitation does contribute positively towards the creativity and novelty of the discussed ideas [16]. However, it would be interesting to look at the differences between domain-specific expert facilitation and just expert facilitation. We could have had a training component to help facilitators brush up on their facilitation skills, but we wanted to avoid the problem of biasing them towards using only one set of strategies. The nature of the task and the way it is structured too has a role to play in the ideation performance [61]. A few participants did mention that they did not possess adequate background knowledge to contribute in a meaningful manner which could have impeded the quality of the ideation conversation. Additionally, our study only focused on observing dyads ideating together which could be radically different from larger crowds. It is still unclear if the ideating group size has an impact on the creative output of the brainstorming session [11, 25, 54]. Since we wanted both the Cue UI and the collaborative note-taking tool to be visible alongside the meeting conference, we dealt with a few space restrictions or challenges to display wordy or very long cues sent out by the facilitator. It is important to be cognizant of this for designing real-time UI-based conversational support for conferencing applications when multiple features are concurrently used. Unfortunately, we encountered a few technical issues with the screen recording tool due to which the recorded reflections for a few cues (around 10), the overall timeliness rating for three ideators in the final post-survey were non-retrievable due to which they had to be removed from our analysis.

Our design probe allowed us to observe the impact of an AI-facilitated virtual ideation discussion. We cover preferences, inconveniences, and wants from both ideators and facilitators based on their experience with the design probe. In our study, we looked into the timeliness of cues being sent out by the facilitators, but we did not delve into the differences in the amount of information and the corresponding impact of each cue. Similarly, our design probe only supported text-based cues that can be sent out by the facilitator but the Generative-AI space has now made it possible to generate multi-modal outputs such as image, video and music. It could be worthwhile looking into ideator preferences for the different modalities of inspiration at different points during brainstorming conversations and the corresponding creative output. We discussed challenges associated with the way the ideators were notified when a new cue was populated and how that could have been a reason as to why ideators missed well-timed and helpful cues. Since timeliness depends on multiple factors such as the content of the cue, awareness regarding if it is repetitive or addressing something that has already been addressed, and many others, future work could delve into experimenting with conversational analysis techniques that can identify what the appropriate content and timing for a support cue should be. Therefore, best practices to present these cues elegantly in real-time collaborative settings is a crucial avenue of research that could inform the design of conferencing tools for creative ideation. Future work could also dive into exploring how Generative-AI can be used to enhance facilitator systems. Maybe this includes helping them with automatic prioritization and quick formulation of useful cues to the ideators. The concept of Gen AI

conversational cues for ideation could go beyond the typical dyad setting, and potentially pervade into facilitating multiple breakout rooms and other large-scale ideation contexts.

7 CONCLUSION

We built a meeting interface where a hidden human wizard facilitator emulated an AI facilitator to send conversational cues to facilitate an ideation discussion between dyads using the Wizard of Oz setup. To grasp an understanding of the general brainstorming facilitation strategies and the respective conversational outcome, we performed a between-subjects study with 14 triads (n=42) who were recruited and asked to participate in an ideation task. Using a thematic analysis, we present insights on conversational cueing strategies from facilitators, their conversational signals of interest, challenges faced by both facilitators and ideators during real-time ideation discussions, and design objectives that are aimed towards targeting these challenges. The findings of this work are intended to guide the development of future speech-driven real-time collaborative ideation support.

ACKNOWLEDGMENTS

We thank the National Science Foundation for funding this project through grant #2009003. Special appreciation is owed to Jane E for her help with proofreading the paper. We are also grateful to the members of the Protolab for their feedback and support. Our heartfelt thanks go to the dedicated members of the Design Lab Operations team for their hard work, which allowed us to focus on our research endeavors. We also thank Vidya Madhavan for assisting with the qualitative analysis. Finally, we extend our thanks to the reviewers for their thorough and constructive feedback, which significantly contributed to the refinement of this work.

Note: Generative AI was used to generate the conversational context for each of the strategy examples (Fig 6, Fig 7, Fig 8, and Fig 9) and diagnose errors for design probe development. It was not used for writing sections of the paper. Image credit for Fig 1(A), 1(B) and 1(C) Source: https://undraw.co/license, Image credit for robot icon: https://icons8.com/icon/CHBf5jmRzl9y/music-robot.

REFERENCES

[1] Joseph A. Allen, Tammy Beck, Cliff W. Scott, and Steven G. Rogelberg. 2014. Understanding workplace meetings: A qualitative taxonomy of meeting purposes. *Management Research Review* 37, 9 (2014), 791–814.

[2] Olaf Adan and Steven Houben. 2023. CollEagle; Tangible Human-AI Interaction for Collocated Collaboration. (2023).

[3] Andrea Agostinelli, Timo I Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. 2023. Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325* (2023).

[4] Salvatore Andolina, Khalil Klouche, Diogo Cabral, Tuukka Ruotsalo, and Giulio Jacucci. 2015. InspirationWall: Supporting Idea Generation Through Automatic Information Exploration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition* (Glasgow, United Kingdom) *(C&C '15)*. Association for Computing Machinery, New York, NY, USA, 103–106. https://doi.org/10.1145/2757226.2757252

[5] Bon Adriel Aseniero, Marios Constantinides, Sagar Joglekar, Ke Zhou, and Daniele Quercia. 2020. MeetCues: Supporting online meetings experience. In *2020 IEEE Visualization Conference (VIS)*. IEEE, 236–240.

[6] Sumit Asthana, Sagih Hilleli, Pengcheng He, and Aaron Halfaker. 2023. Summaries, Highlights, and Action items: Design, implementation and evaluation of an LLM-powered meeting recap system. arXiv:2307.15793 [cs.HC]

[7] Patti Bao, Elizabeth Gerber, Darren Gergle, and David Hoffman. 2010. Momentum: getting and staying on topic during a brainstorm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1233–1236.

[8] Tony Bergstrom and Karrie Karahalios. 2007. Conversation Clock: Visualizing audio patterns in co-located groups. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. IEEE, 78–78.

[9] Tony Bergstrom and Karrie Karahalios. 2009. Conversation clusters: grouping conversation topics through human-computer dialog. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2349–2352.

[10] Jennie Björk. 2012. Knowledge domain spanners in ideation. *Creativity and innovation management* 21, 1 (2012), 17–27.

[11] Thomas J Bouchard Jr and Melana Hare. 1970. Size, performance, and potential in brainstorming groups. *Journal of applied Psychology* 54, 1p1 (1970), 51.

[12] John H Bradley and Frederic J Hebert. 1997. The effect of personality type on team performance. *Journal of Management Development* 16, 5 (1997), 337–353.

[13] Francisco Maria Calisto, João Fernandes, Margarida Morais, Carlos Santiago, João Maria Abrantes, Nuno Nunes, and Jacinto C Nascimento. 2023. Assertiveness-based agent communication for a personalized medicine on medical imaging diagnosis. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–20.

[14] Joana Campos, James Kennedy, and Jill F Lehman. 2018. Challenges in exploiting conversational memory in human-agent interaction. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 1649–1657.

[15] Joel Chan, Steven Dang, and Steven P Dow. 2016. IdeaGens: enabling expert facilitation of crowd brainstorming. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. 13–16.

[16] Joel Chan, Steven Dang, and Steven P Dow. 2016. Improving crowd innovation with expert facilitation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1223–1235.

[17] 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.

[18] Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang, and Kwan-Liu Ma. 2019. TalkTraces: Real-time capture and visualization of verbal content in meetings. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

[19] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

[20] Carsten KW De Dreu and Laurie R Weingart. 2003. Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis. *Journal of applied Psychology* 88, 4 (2003), 741.

[21] Michael Diehl and Wolfgang Stroebe. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of personality and social psychology* 53, 3 (1987), 497.

[22] Hyo Jin Do, Ha-Kyung Kong, Pooja Tetali, Jaewook Lee, and Brian P Bailey. 2023. To Err is AI: Imperfect Interventions and Repair in a Conversational Agent Facilitating Group Chat Discussions. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–23.

[23] Steven Dow, Blair MacIntyre, Jaemin Lee, Christopher Oezbek, Jay David Bolter, and Maribeth Gandy. 2005. Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing* 4, 4 (2005), 18–26.

[24] Steven P Dow, Manish Mehta, Blair MacIntyre, and Michael Mateas. 2010. Eliza meets the wizard-of-oz: blending machine and human control of embodied characters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 547–556.

[25] R Brent Gallupe, Alan R Dennis, William H Cooper, Joseph S Valacich, Lana M Bastianutti, and Jay F Nunamaker Jr. 1992. Electronic brainstorming and group size. *Academy of Management Journal* 35, 2 (1992), 350–369.

[26] Diether Gebert, Sabine Boerner, and Eric Kearney. 2010. Fostering team innovation: Why is it important to combine opposing action strategies? *Organization Science* 21, 3 (2010), 593–608.

[27] Amr Gomaa and Michael Feld. 2023. Towards Adaptive User-centered Neuro-symbolic Learning for Multimodal Interaction with Autonomous Systems. In *Proceedings of the 25th International Conference on Multimodal Interaction*. 689–694.

[28] Joy P Guilford. 1967. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior* 1, 1 (1967), 3–14.

[29] John P Guilford. 1984. Varieties of divergent production. *The Journal of Creative Behavior* (1984).

[30] Eleazar Hernández and Eleazar Hernández. 2017. Brainstorming: Where do all those crazy ideas come from? *Leading Creative Teams: Management Career Paths for Designers, Developers, and Copywriters* (2017), 57–74.

[31] Cindy E Hmelo-Silver and Howard S Barrows. 2006. Goals and strategies of a problem-based learning facilitator. *Interdisciplinary journal of problem-based learning* 1, 1 (2006), 4.

[32] Thomas J Howard, Elies A Dekoninck, and Steve J Culley. 2010. The use of creative stimuli at early stages of industrial product innovation. *Research in Engineering design* 21 (2010), 263–274.

[33] Bernd Huber, Stuart Shieber, and Krzysztof Z Gajos. 2019. Automatically analyzing brainstorming language behavior with Meeter. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–17.

[34] Irving L Janis. 1971. Groupthink: The desperate drive for consensus at any cost. *Classics of organization theory* 6 (1971), 185–192.

[35] Peiling Jiang, Jude Rayan, Steven P Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. *arXiv preprint arXiv:2305.11473* (2023).

[36] Vaiva Kalnikaitė, Patrick Ehlen, and Steve Whittaker. 2012. Markup as you talk: establishing effective memory cues while still contributing to a meeting. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. 349–358.

[37] Reet Kasepalu, Luis P Prieto, Tobias Ley, and Pankaj Chejara. 2022. Teacher artificial intelligence-supported pedagogical actions in collaborative learning coregulation: A wizard-of-oz study. In *Frontiers in Education*, Vol. 7. Frontiers, 736194.

[38] Tae Soo Kim, Seungsu Kim, Yoonseo Choi, and Juho Kim. 2021. Winder: linking speech and visual objects to support communication in asynchronous collaboration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[39] Stefan Werner Knoll and Graham Horton. 2010. Changing the perspective: improving generate thinkLets for ideation. In *2010 43rd Hawaii International Conference on System Sciences*. IEEE, 1–10.

[40] Nicholas W Kohn and Steven M Smith. 2011. Collaborative fixation: Effects of others' ideas on brainstorming. *Applied Cognitive Psychology* 25, 3 (2011), 359–371.

[41] JiayiZhou. Renzhong Li, Junxiu Tang, Tan Tang, Haotian Li, Weiwei Cui, and Yingcai Wu. 2024. Understanding Nonlinear Collaboration between Human and AI Agents: A Co-design Framework for Creative Design. arXiv:2401.07312 [cs.HC]

[42] Xingyu" Bruce" Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Alex Olwal, Peggy Chi, Xiang" Anthony" Chen, and Ruofei Du. 2023. Visual Captions: Augmenting Verbal Communication With On-the-Fly Visuals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[43] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J Cai. 2020. Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[44] Kent Lyons, Christopher Skeels, Thad Starner, Cornelis M Snoeck, Benjamin A Wong, and Daniel Ashbrook. 2004. Augmenting conversations using dual-purpose speech. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*. 237–246.

[45] O. Miksik, I. Munasinghe, J. Asensio-Cubero, S. Reddy Bethi, S-T. Huang, S. Zylfo, X. Liu, T. Nica, A. Mitrocsak, S. Mezza, R. Beard, R. Shi, R. Ng, P. Mediano, Z. Fountas, S-H. Lee, J. Medvesek, H. Zhuang, Y. Rogers, and P. Swietojanski. 2020. Building Proactive Voice Assistants: When and How (not) to Interact. arXiv:2005.01322 [cs.HC]

[46] Elinor Mizrahi, Noa Danzig, and Goren Gordon. 2022. vRobotator: A virtual robot facilitator of small group discussions for K-12. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–22.

[47] Lennart Molin. 2004. Wizard-of-Oz prototyping for co-operative interaction design of graphical user interfaces. In *Proceedings of the third Nordic conference on Human-computer interaction*. 425–428.

[48] Hazel Morton and Mervyn A Jack. 2005. Scenario-based spoken interaction with virtual agents. *Computer Assisted Language Learning* 18, 3 (2005), 171–191.

[49] Bernard A Nijstad, Carsten KW De Dreu, Eric F Rietzschel, and Matthijs Baas. 2010. The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European review of social psychology* 21, 1 (2010), 34–77.

[50] Chunjong Park, Junsung Lim, Juho Kim, Sung-Ju Lee, and Dongman Lee. 2017. Don't bother me. I'm socializing! A breakpoint-based smartphone notification system. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 541–554.

[51] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.

[52] Sidney J Parnes. 1961. Effects of extended effort in creative problem solving. *Journal of Educational psychology* 52, 3 (1961), 117.

[53] Paul Paulus. 2000. Groups, teams, and creativity: The creative potential of idea-generating groups. *Applied psychology* 49, 2 (2000), 237–262.

[54] Paul B Paulus, Nicholas W Kohn, Lauren E Arditti, and Runa M Korde. 2013. Understanding the group size effect in electronic brainstorming. *Small Group Research* 44, 3 (2013), 332–352.

[55] Paul B Paulus, Vicky L Putman, Karen Leggett Dugosh, Mary T Dzindolet, and Hamit Coskun. 2002. Social and cognitive influences in group brainstorming: Predicting production gains and losses. *European review of social psychology* 12, 1 (2002), 299–325.

[56] Baolin Peng, Michel Galley, Pengcheng He, Hao Cheng, Yujia Xie, Yu Hu, Qiuyuan Huang, Lars Liden, Zhou Yu, Weizhu Chen, and Jianfeng Gao. 2023. Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback. arXiv:2302.12813 [cs.CL]

[57] Anna Penzkofer, Philipp Müller, Felix Bühler, Sven Mayer, and Andreas Bulling. 2021. Conan: A usable tool for multimodal conversation analysis. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 341–351.

[58] Nia Peters, Griffin Romigh, George Bradley, and Bhiksha Raj. 2017. When to interrupt: A comparative analysis of interruption timings within collaborative communication tasks. In *Advances in Human Factors and System Interactions: Proceedings of the AHFE 2016 International Conference on Human Factors and System Interactions, July 27-31, 2016, Walt Disney World®, Florida, USA*. Springer, 177–187.

[59] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. Anglekindling: Supporting journalistic angle ideation with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[60] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv:2204.06125 [cs.CV]

[61] Eric F Rietzschel, J Marjette Slijkhuis, and Nico W Van Yperen. 2014. Task structure, need for structure, and creativity. *European Journal of Social Psychology* 44, 4 (2014), 386–399.

[62] Simone M Ritter and Nel 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 (2018), 263–277.

[63] ML Runway. 2023. Advancing creativity with artificial intelligence.

[64] Samiha Samrose, Daniel McDuff, Robert Sim, Jina Suh, Kael Rowan, Javier Hernandez, Sean Rintel, Kevin Moynihan, and Mary Czerwinski. 2021. Meetingcoach: An intelligent dashboard for supporting effective & inclusive meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[65] Elizabeth B-N Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *Co-design* 4, 1 (2008), 5–18.

[66] Arissa J Sato, Zefan Sramek, and Koji Yatani. 2023. Groupnamics: Designing an Interface for Overviewing and Managing Parallel Group Discussions in an Online Classroom. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[67] Yang Shi, Chris Bryan, Sridatt Bhamidipati, Ying Zhao, Yaoxue Zhang, and Kwan-Liu Ma. 2018. Meetingvis: Visual narratives to assist in recalling meeting context and content. *IEEE Transactions on Visualization and Computer Graphics* 24, 6 (2018), 1918–1929.

[68] Pao Siangliulue, Joel Chan, Steven P Dow, and Krzysztof Z Gajos. 2016. IdeaHound: improving large-scale collaborative ideation with crowd-powered real-time semantic modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 609–624.

[69] James Simpson, Hamish Stening, Patrick Nalepka, Mark Dras, Erik D Reichle, Simon Hosking, Christopher J Best, Deborah Richards, and Michael J Richardson. 2022. DesertWoZ: A Wizard of Oz Environment to Support the Design of Collaborative Conversational Agents. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*. 188–192.

[70] Donggil Song, Marilyn Rice, and Eun Young Oh. 2019. Participation in online courses and interaction with a virtual agent. *International Review of Research in Open and Distributed Learning* 20, 1 (2019).

[71] Ben Swanson, Kory Mathewson, Ben Pietrzak, Sherol Chen, and Monica Dinalescu. 2021. Story centaur: Large language model few shot learning as a creative writing tool. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. 244–256.

[72] Sander Välk, Chitipat Thabsuwan, and Céline Mougenot. 2023. The Ideation Compass: supporting interdisciplinary creative dialogues with real time visualization. *International Journal of Design Creativity and Innovation* 11, 2 (2023), 99–116.

[73] Hao-Chuan Wang, Dan Cosley, and Susan R Fussell. 2010. Idea expander: supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 103–106.

[74] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B Chilton. 2023. PopBlends: Strategies for conceptual blending with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[75] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–22.

[76] Haijun Xia, Tony Wang, Aditya Gunturu, Peiling Jiang, William Duan, and Xiaoshuo Yao. 2023. CrossTalk: Intelligent Substrates for Language-Oriented Interaction in Video-Based Communication and Collaboration. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[77] Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, Longyue Wang, Anh Tuan Luu, Wei Bi, Freda Shi, and Shuming Shi. 2023. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. arXiv:2309.01219 [cs.CL]

[78] Qingxiao Zheng, Yiliu Tang, Yiren Liu, Weizi Liu, and Yun Huang. 2022. UX research on conversational human-AI interaction: A literature review of the ACM digital library. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–24.