Group Brainstorming with an AI Agent: Creating and Selecting Ideas

Michael Muller[1], **Stephanie Houde**[1], **Gabriel Gonzalez**[2], **Kristina Brimijoin**[1],
Steven I. Ross[1], **Dario Andres Silva Moran**[2], **Justin D. Weisz**[1]
[1]IBM Research AI, USA; [2]IBM Argentina, AR
{michael_muller, kbrimij, steven_ross, jweisz}@us.ibm.com
{stephanie.houde, gabriel.gonzalez, dario.silva}@ibm.com

Abstract

Researchers have experimented with ways of providing computer assistance to the co-creative task of brainstorming. Now, large language models (LLMs) present new opportunities and challenges to bring an AI agent directly into a brainstorming session.

We built an AI agent to act as an interactive participant in online conversational brainstorming for a distributed workforce. Eighteen colleagues participated in 6 brainstorming experiences (3 people per replication, 3 topics across 3 sessions, counterbalanced) with an AI as a "fourth participant." At the end of each session, participants chose 3 ideas as "final" i.e., to be recommended to an imagined client. Humans and AI collaborated in creating, evaluating, refining, and selecting a larger number ideas through five different patterns of idea-development.

Using frameworks from mixed initiative interfaces, we analyze five types of actions taken by humans and by AI, and we begin to answer the research question: How does an idea become final?

Introduction

Brainstorming is a familiar activity in everyday co-creativity. (Paulus and Brown 2003). Researchers have explored computer-assisted e-brainstorming between human participants (Liikkanen et al. 2011), but now large language models (LLMs) present new opportunities and challenges to bring an AI agent directly into a brainstorming session (Bouschery et al. 2022).

Building on previous research that involved one human and one AI (Bouschery et al. 2022; Muller, Candello, and Weisz 2023), we designed an experiment in which a group of humans engaged with an AI agent. We built an AI agent as a participant in online brainstorming for a distributed workforce (Orit Shaer and Mokryn 2024). Eighteen colleagues participated in 6 conversational brainstorming experiences in Slack [1] (3 people per replication, 3 topics across 3 sessions, counterbalanced). In 2 of each of the 3 sessions, the AI was present as a "fourth participant." At the end of each session, participants chose 3 ideas as "final" i.e., to be recommended to an imaginary client as the 3 best ideas from the session.

[1]https://slack.com.

We perform an analysis of the sequences of human and AI actions that led to the selection of those final ideas. Using frameworks from mixed initiative interfaces (Grabe, Duque, and Zhu 2022; Muller, Weisz, and Geyer 2020; Spoto and Oleynik 2017), we analyzed five types of actions taken by humans and by AI, and we begin to answer the question: *How does an idea become final?*

Related Work

Brainstorming as everyday co-creativity

Humans have engaged in co-creative tasks similar to brainstorming for centuries (Paulus and Brown 2003). In the past several decades, researchers have experimented with computer-based aids to brainstorming, sometimes called e-brainstorming (Liikkanen et al. 2011; Paulus 2015; Mongeau and Morr 1999). These technological approaches allow humans to parallelize their initial idea-generation, removing theorized obstacles to brainstorming such as production blocking, in which people have to wait to state their own ideas until others have finished speaking (Nijstad, Stroebe, and Lodewijkx 2003; Stroebe, Nijstad, and Rietzschel 2010).

Beyond the initial production of ideas, researchers report numerous benefits of group brainstorming, including "cognitive stimulation" (Dugosh et al. 2000) through learning of other brainstormers' ideas (Fink et al. 2012; Paulus and Dzindolet 2008; Santanen, Briggs, and De Vreede 2000) and fertile combinations of ideas from multiple perspectives (John-Steiner 2000; Maher 2012; Suh et al. 2021). These advantages may be particularly important for teams and others who are building collective competence while they are co-creating ideas (Paulus, Dzindolet, and Kohn 2012).

AI and Brainstorming

Frameworks. In this project, we added an AI agent to groups of humans engaged in brainstorming - an everyday workplace form of co-creativity. We wanted to study what happened when an AI agent took roles similar to those performed by humans. We follow insights from Glăveanu's *Distributed creativity* (Glăveanu 2014), which applied more general principles from studies of *Distributed cognition* (Hollan, Hutchins, and Kirsh 2000). In these approaches, cognition and/or creativity take place among multiple humans, and may depend on forms of knowledge or intelligence that are stored or implemented in objects and/or computational entities. Importantly, these research traditions go beyond the individual knower or actor, to describe groups of knowers and actors, who may partially share knowledge and actions with non-human entities.

More specifically, we build on Kantosalo's and Takala's 5Cs framework, which describes "a human-computer *collective*, the collective's *collaboration* process and creative *contributions* to a *community* ... within a rich *context*" (Kantosalo and Takala 2020, p. 17, italics in the original text). We broaden the collective from *one* person working with an AI agent, to a *group* of people working with that AI agent, and we focus on the collaboration process through an analysis of the structure of their brainstorming activities.

Brainstorming is a condition mid-way between Maher et al.'s *collected* intelligence vs. *collective* intelligence (2011), because some ideas are generated by individual humans or AI agents, and other ideas are discussed and supported by multiple entities. In Rezwana's and Maher's (2021) COFI taxonomy, our Task Distribution was *Same-task*; our bidirectional Communications media were *Text*; our Creative Process was *Generate, Evaluate*; and our Contribution Types were *Create-new, Extend, Refine, and Transform*.

Technology-Centric Approaches. Previous work has examined human-AI co-creativity in ideation tasks. Koch et al. (2019) experimented with an AI agent that could add content to a moodboard controlled by a single human. Oh et al. (2018) took a more directly engaged approach through collaborative drawing among a single human and an AI. Work with text-to-image systems has also informed collaborations between one human and an AI (Chiou et al. 2023; Millwood and Dias-Taguatinga 2023). Farrell et al. proposed an architecture of cognitive agents (a "society of cogs") that could interact with one or two humans (Farrell et al. 2016).

Human-Centric Approaches. The preceding paragraph described technology-centric methods for human-AI creative collaborations. We now turn to more human-centric approaches. Several groups examined the effect of telling users that they were working with either an AI or another human being, with contradictory results (Geerts, de Wit, and de Rooij 2021; Wieland, de Wit, and de Rooij 2022; Yu-Han and Chun-Ching 2023). Lavrič and Škraba (2023) experimented with AI-generated materials that were subsequently discussed among groups of humans.

Three other sets of papers come closer to our project, in that they involved a human and an AI in directly shared ideation tasks. Muller et al. described plausible design ideation and brainstorming in informal experiments between one human and an AI (Muller, Candello, and Weisz 2023; Muller 2023; Muller and Weisz 2023). Bouschery et al. conducted a more formal comparison of isolated or paired human-AI configurations, reporting superior results with a human and an AI working together (Bouschery, Blazevic, and Piller 2023). Shaer et al. study structured group "Brainwriting" that incorporates AI generated content at key points in a group ideation process (Shaer et al. 2024).

Table 1: **Mixed Initiative Action Vocabularies**

MICI[a]	**MIGAI**[b]	**FashionGAN**[c]	**Koala**[d]
		Initialize	
	Learn	Learn	
Ideate	Ideate		Idea
Constrain	Constrain	Constrain	Task instructions
Produce	Produce	Create	
Suggest	Suggest	Select	
Select	Select	Select	
	Curate		Final
Assess	Assess		Plus & Minus
Adapt	Adapt	Adapt	Refine
	Assemble	Combine	
	Wait		

Notes:
a. Spoto and Oleynik (2017). b. Muller, Weisz, and Geyer (2020). c. Grabe, Duque, and Zhu (2022). d. Brainstorming actions referenced in this paper.

This paper expands on the previous work by including a *group* of humans, working on a shared brainstorming task, with *direct interactive* involvement by an AI agent. We examine in detail the roles of humans and AI, and we detail the distinct activities that humans and AI performed in the course of active brainstorming.

Mixed Initiative Action Vocabularies

To analyze human and AI activities, we drew from work in mixed initiative creative interfaces (Horvitz 1999). Spoto and Oleynik (2017) developed a 7-action vocabulary for mixed initiative interfaces, based on creative applications from a CHI 2017 workshop (Deterding et al. 2017). A key insight was that either human or AI could perform each of those actions, and that each application had its own distinctive *sequence* of actions. Two variations on the Spoto and Oleynik vocabularies appeared subsequently, as shown in Table 1, with further development of distinctive or characteristic patterns of actions (Grabe, Duque, and Zhu 2022; Muller, Weisz, and Geyer 2020). For this paper, we mapped selected actions into the brainstorm use-case. We will present this subset of actions in detail, in the section *Defining a Sequence of Events*.

Method

Koala

We developed Koala, a multi-party conversational AI agent that can collaborate with humans in co-creative tasks. Koala was implemented with a backend powered by a Llama 2 large language model (Touvron et al. 2023). The agent communicated via the Slack API[2] so that it could act as a slackbot (Laitinen, Laaksonen, and Koivula 2021) and operate as a participant in a Slack channel (Seering et al. 2019). Figure 1 shows an example of Koala's presence in Slack. To avoid ethical issues with AI anthropomorphism

[2] https://api.slack.com

(Kuss and Leenes 2020; Salles, Evers, and Farisco 2020; Shneiderman and Muller 2023), we clearly labeled Koala as an AI agent via an "APP" suffix in each conversational turn.

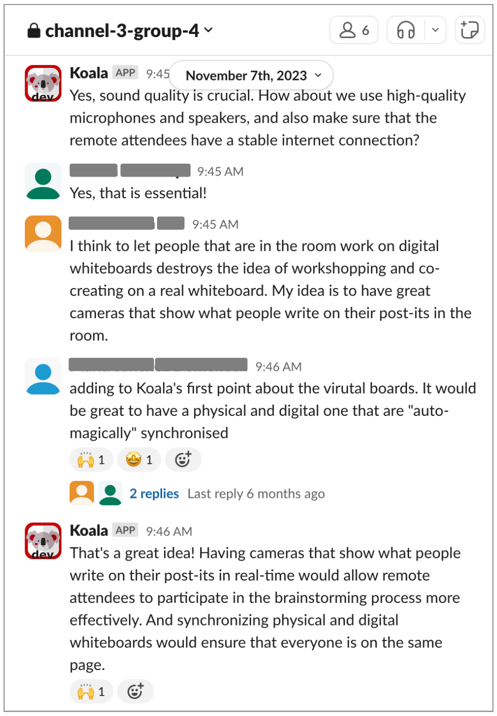

Figure 1: **Excerpt from Session 4 Brainstorming** Three humans and one Proactive AI (Koala) discussed ideas in a Slack channel to address the topic: *How can we improve Hybrid meetings for remote participants?*

Participants

We recruited 18 participants via Slack channels within a large multinational technology corporation. They came from 3 countries on 3 continents (10 women, 7 men, 1 prefer-not-to-answer). Participants' job roles included design, user research, videography, and software engineering. All participants had some familiarity with AI-based products and services. Each participant received the equivalent of $US25.00 as compensation for one hour of participation.

Procedure

Six groups of 3 participants each engaged in 3 brainstorming sessions. One baseline session had no AI agent. Two other sessions had one of two versions of the Koala agent: Reactive AI, where Koala participated only when asked to, and Proactive AI, where Koala could participate proactively by deciding on its own whether to contribute to the conversation as well as on-demand. In this paper, we focus on human-human-AI interactions in the two AI-present conditions.

In each session, participants worked on a different brainstorming topic. We chose topics on the basis of generality and familiarity, so that all humans would have a common starting point:

- How can we improve hybrid meetings for remote participants?
- How can we better engage employees to use a chatbot for HR services?
- What kinds of give-aways should we take to a customer conference?

Topics were counterbalanced across the 6 groups, such that each topic occurred equally often in the first, second, and third session, and each topic occurred equally often with each of the versions of the AI. Figure 1 shows an except from a brainstorming discussion in session 4, where three human participants and the Proactive version of Koala discuss ideas to address the hybrid meetings topic.

Over the course of the study, we collected transcripts of the brainstorming sessions and survey data from the participants. This paper focuses on the transcript data only, tracking the actions that humans and AIs took while discussing the ideas that humans and AI proposed in the brainstorming tasks.

Research Question

The crucial outcome of each session was the three recommended final ideas. Therefore, we focus on this **RQ: How does an idea become a final idea?** To answer that question, we explored which factors, and which sequences of actions, were associated with an idea being selected as final.

Results

Survey data collected during the study indicated that 100% of participants preferred AI-supported brainstorming to the no AI condition, and that the AI conditions led to significantly more ideas. Participants in each session were instructed to select 3 ideas as "final."

In this paper we focus on granular analysis of the conversational transcripts to study the sequence of actions that each group of 3 humans and 1 AI performed in order to arrive at their selection of final ideas.

Understanding Collaborative Ideation

We began our analysis by visualizing, for each session, the relationships of Humans and AIs to the ideas that they proposed. There was no "typical" session. Figures 2 and 3 show two approaches to understand session-to-session commonalities and differences.

Figure 2 shows an example from Session 1, in which humans (*H1*, *H2*, *H3*) and the AI agent (*AI*) are linked to the ideas that they proposed or discussed. Gold diamonds indicate ideas that humans selected as final ideas. This visualization shows the AI engaging in almost all the ideas in one capacity or another, while the human participants seem to be restricted to engaging with 1, 2, or 4 of the 14 ideas listed.

We created a second set of visualizations of the structure of discussions within each session. Figure 3 provides an example for the same session as in Figure 2. This second visualization emphasizes the relationships among conversational (brainstorming) turns and the ideas that are referred to in those conversational turns. The participants all engaged

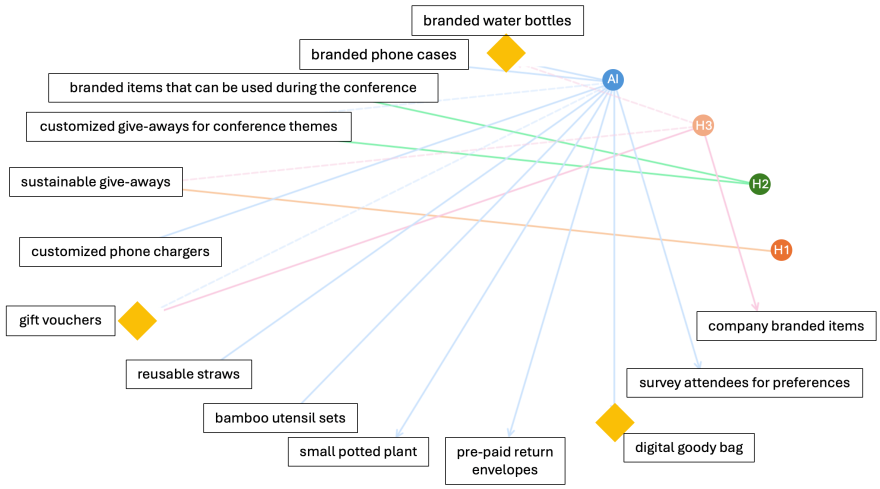

Figure 2: **Mapping of human (*H1*, *H2*, *H3*) and AI agent (*AI*) to the ideas proposed and discussed in Session 1.** Solid lines indicate idea creation. Dotted lines indicate idea refinement. Gold diamonds indicate the three ideas that humans selected as "final" - i.e., as outcomes that they would recommend to a client.

with each other and with the AI, and the AI engaged with the other participants.

Of the three final ideas selected, one was originally suggested by the AI, and one was a refinement of an idea suggested by the AI, so the AI clearly made an valued contribution to the final outcome. The other final idea was refined by the AI, but the refinement wasn't selected, and the original idea also had objections raised by a human participant. The AI was the only participant suggesting multiple ideas on a single turn, which it did twice. For roughly the second half of the conversation (by turn, not necessarily by time) the AI was the only one continuing to suggest new ideas, while the other participants were mostly just refining and reacting to ideas that had been suggested earlier.

Defining a Sequence of Events for each Idea

Based on ideas from the visualizations, we begin statistical analyses of the "lineage" of each idea as it progresses from the moment in which it is proposed (by human or by AI) to the moment at which it is selected as one of the 3 "final" ideas, or is not selected ("nonfinal").

For each idea, we coded the text of the transcripts to represent the Sequence of events from the initial proposal of the idea to its selection as final or nonfinal. Each event in the sequence was defined as ***Ea***. The first letter, ***E***, represents the Entity that performs that action - either ***H***uman or ***A***I. The second letter, ***a***, represents the Action that the Entity took - one of ***i***dea (propose the idea); ***p***lus (make a statement in support of the idea); ***m***inus (make a statement against the idea); ***r***efine (suggest concepts or further develop the idea); and ***f***inal (select the idea as one of the 3 final ideas). Three of the authors coded all the transcripts. Coders resolved any disagreements in codes via discussion.

Human and AI Events

First, we examined whether Humans and AI used the five Actions with the same frequencies. Figure 4A shows the simple counts of Human vs. AI actions. These frequency distributions of Actions were significantly different for Humans vs. AI, with $\chi^2_{(4)}$ = 118.861, p $<$ 0.001. Humans used more plus actions, while the AI initiated more ideas. In Figure 4B, we present the data adjusted to remove the confound that there are 3 humans and only one AI. The differences observed in Figure 4A are even more pronounced in Figure 4B, with $\chi^2_{(4)}$ = 72.476, p $<$ 0.001.

Figures 4C and 4D split the data into two subsets: One for humans, and one for AI. In Figure 4C, humans used disproportionately more plus actions for final ideas, with $\chi^2_{(3)}$ = 44.573, p $<$ 0.001. In Figure 4D, the AI proposed disproportionately more ideas that did not become final, and used disproportionately more plus actions on those ideas, with $\chi^2_{(3)}$ = 49.294, p $<$ 0.001. (We note that we omitted the final actions from both final-vs.-nonfinal analyses, because nonfinal ideas had zero final actions by definition, and would have artificially inflated the chi-square statistic).

Humans and AI showed different usage patterns among the five Actions. The AI proposed many more ideas than any single human, and most of those ideas became nonfinal. Humans supported (plus-action) more ideas that became final vs. nonfinal, while the pattern for the AI was reversed, with more support (plus) actions on ideas that did not become final.

Humans choose AI-initiated ideas as final in 33% of the cases. Thereby, humans did not reject all AI ideas. Nonetheless, our results show that AI-initiated ideas were on average less likely to be accepted as final, and that the AI supported more ideas that did not become final.

Becoming Final. Thereby, we begin to answer our question about how an idea becomes final. Human-proposed

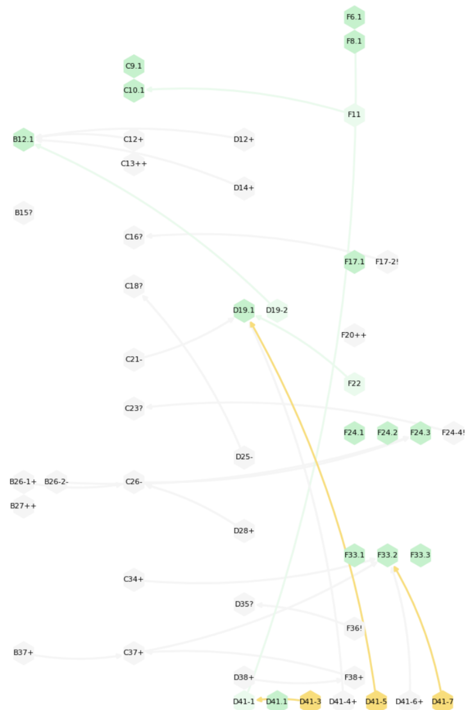

Figure 3: **Structure of the brainstorming discussion in Session 1.** The timeline reads from top to bottom. Green hexagons represent new ideas and light green hexagons represent refinements of ideas. Grey hexagons represent comments on ideas, with a suffix ("+", "-", "?", "!") to indicate a positive or "plus" comment, a negative or "minus" comment, a question, or an answer respectively. The three gold hexagons represent the selection of each of the final ideas. Humans are indicated as "B", "C", "D", while AI is indicated as "F." Directed arcs show the relationship of each conversational turn to a specific idea - e.g., at the bottom of the figure, D41.7 is an action by Human D to choose AI's idea F33.2 as a final idea.

ideas are more likely to become final. AI-proposed ideas may also become final, but with a lower probability.

However, we have been writing as if the AI's ideas and the humans' ideas were separate from each other. In the next subsections, we will show that this is not true - i.e., that humans and AI collaborated on most ideas that were selected as final.

Sequential Analyses of Entities

Figure 5 shows a Sankey diagram[3] of 171 Sequences of actions, from the initial proposal of an idea by Human (*Hi*) or by AI (*Ai*) at the left side of the Figure, to the Outcome (final or nonfinal) at the right side of the Figure. Sankey diagrams are typically used to track resources, finances, or persons who enter a process or workflow, are acted upon during that process, and then exit the process or workflow into different categories (Franceys et al. 2017; Hernandez et al. 2018; Lamer et al. 2020). In our case, each item is an idea, and the process is the sequence of actions that humans and AI perform in relation to each idea. The Sankey diagram suggests several important points for statistical testing.

First, many AI-proposed ideas (Ai, n=124 ideas, 73% of all ideas) were never discussed by humans or by AI (n=84, 68% of AI-proposed ideas). A smaller number of human-proposed ideas (Hi, n=47 ideas, 27% of all ideas) were similarly ignored by the AI (n=17, 36% of human-proposed ideas). In total, 82 ideas (48% of all ideas) did not receive attention from either human or AI. Not all ideas are valuable.

Second, there was no single, characteristic series of events that led to either final or nonfinal. Figure 5 shows many different paths from idea-proposal to final outcomes and to nonfinal outcomes. We examine some factors that make partial contributions to final/nonfinal outcomes, in the remainder of this section.

There was a tendency for final ideas to be decided early (e.g., 69% ideas by step 7), while some nonfinal ideas were discussed through many more steps (maximum of 14 steps). The differences in Sequence lengths between final and nonfinal was significant at $F_{(1,169)} = 134.052$, $p < 0.001$.

Becoming Final. So far, we have learned that action on an idea is more likely to lead to that idea becoming final. This observation is not surprising. What is more interesting is that extended action on an idea is less likely to result in the idea becoming final. Final decisions tend to be made after relatively few conversational exchanges, whereas nonfinal outcomes can involve a longer series of exchanges.

Collaborations

We now examine patterns of collaborations among humans and AI. We begin by restating the baseline of 171 ideas. 82 of those ideas (48%) received no further attention by humans or by AI. Only 3 ideas became final without further action by humans or AI (i.e., from *Hi* or directly to *Af*). The other 33 final ideas required at least one human action or one AI action before they were selected as final. 28 ideas (78% of all 36 final ideas) required both a human action and an AI action.

Human-AI Collaboration. In contrast, 39 nonfinal ideas involved both a human and an AI action (29%). Thus, involvement by at least one human and the AI more than doubled the likelihood for an idea to become final, with $\chi^2_{(1)} = 8.847$, $p < 0.003$. The co-involvement of at least one human and the AI was crucial, with only 10 ideas becoming final by AI actions alone, and only 4 ideas becoming final by human actions alone.

Human-Human Collaborations. We performed the analogous statistics for human-human collaborations (i.e., at least two different humans acting on the same idea). More than two-thirds of the final ideas (n=25, 69%) involved at least one instance of human-human collaboration. Only 13 of the nonfinal ideas (10%) similarly involved human-human collaborations. These differences were significant, with $\chi^2_{(1)} = 58.832$, $p < 0.001$.

[3] https://www.sankeymatic.com/build/

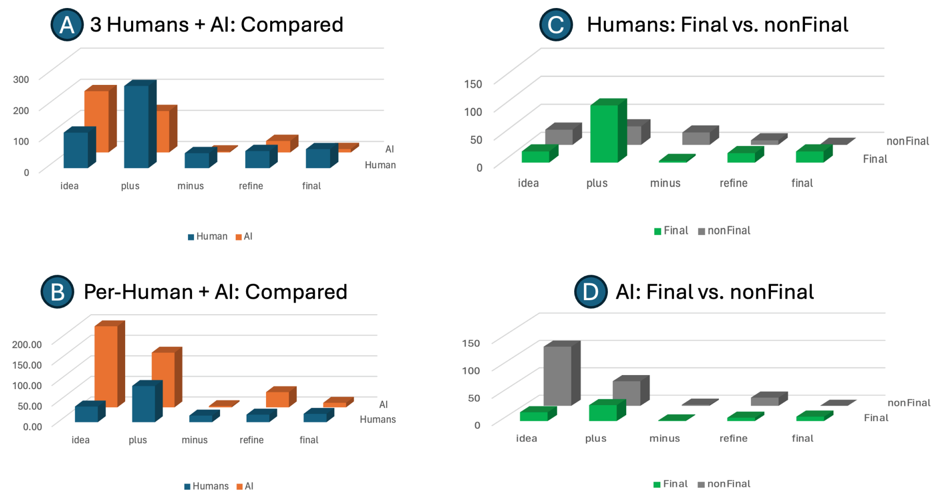

Figure 4: **The distribution of Actions across the two Entities (human and AI).** A. Comparing the simple counts of Actions by each type of Entity. B. Data of part A, adjusted for the number of humans (3) vs. the number of AIs (1). C. Human Actions on final vs. nonfinal ideas. D. AI Actions on final vs. nonfinal ideas.

However, there were six possible combinations of humans, so we also computed statistics after adjusting for this confound (i.e., we divided the human-human collaboration statistics by 6). The differences remained significant, with $\chi^2_{(1)}$ = 20.629, p $<$ 0.001, verified with the approximate Fisher test at $<$ .04.

Becoming Final. Shared actions by at least one human and the AI - or by two humans together - strongly increased the chance for an idea to become final.

Solo Work

The opposite case to collaborations is solo effort by a single human or AI. Possible patterns could include: propose an idea and then support one's own idea; support an idea twice in-a-row; or support an idea and then propose it as final.

We examined repeated individual actors (human or AI) in runs of length=2 - i.e., if the three humans in a session were coded as B, C, D, and if the AI were coded as A, then we looked for sequences of BB, CC, DD, and also AA. For final ideas, the number of human runs was 24, as contrasted with 9 nonfinal ideas. However, we note that there were three opportunities for a human run (BB, CC, and DD) vs. a single opportunity for an AI run (AA). For a fair comparison, we adjust the number of human runs by dividing by 3. The numbers of runs for AI were 12 and 126 for final and nonfinal, respectively.

There were more human runs for final ideas, and more AI runs for nonfinal ideas. This pattern was significant with $\chi^2_{(1)}$ = 66.696, p $<$ 0.001 (unadjusted) and $\chi^2_{(1)}$ = 21.899, p $<$ 0.001 (adjusted).

We wanted to know if final ideas involved combinations of collaborative work and solo work (Figure 6). We counted 30 of 36 final ideas that involved human-AI collaboration (83%), 23 final ideas that involved at least two repeated solo actions by at least one human or AI (64%), and 20 final ideas that involved at least one instance of human-AI collaboration and at least one repeated solo action (56%). While there was not complete overlap of the two phenomena, we observed an average overlap of 76% of final ideas with both phenomena, from the pool of final ideas with either one or the other (or both) phenomena.

Becoming Final. Although the preceding subsection showed that final ideas were associated with collaboration among human(s) and AI, this subsection showed that final ideas were also associated with repeated actions by a single human or by the AI. In just over 50% of the idea sequences, both collaboration and solo work were present.

Sequential Analyses of Actions

We turn now to the actions (e.g., plus, minus, refine) that might influence how an idea becomes final.

First, we performed the simple comparison of final ideas with any of those actions (n=33, 92%) vs. nonfinal ideas with any of those actions (n=53, 39%). The presence of any of those three actions was significantly associated with final outcomes, with $\chi^2_{(1)}$ = 33.668, p $<$ 0.001, verified to be significant at $<$ 0.001 via an approximate Fisher test.

Next, we inquired about outcomes associated with specific actions. Plus actions were significantly associated with final outcomes, with $\chi^2_{(1)}$ = 36.480, p $<$ 0.001, and 89% of final actions including at least one Plus event in their sequence. There were no significant results for Minus actions, with $\chi^2_{(1)}$ = 0.214, n.s. Surprisingly, there was an inverse relationship for Refine actions, with $\chi^2_{(1)}$ = 6.697, p $<$ .02, and 69% of nonfinal actions including at least one Refine action in their sequence. It may be that a Refine event is often a way to discuss weaknesses in the idea.

Which of the three *specific* actions was associated with final outcomes? Because they were the most frequently used actions, we tested for *plus* actions (*Hp* and *Ap*). We found

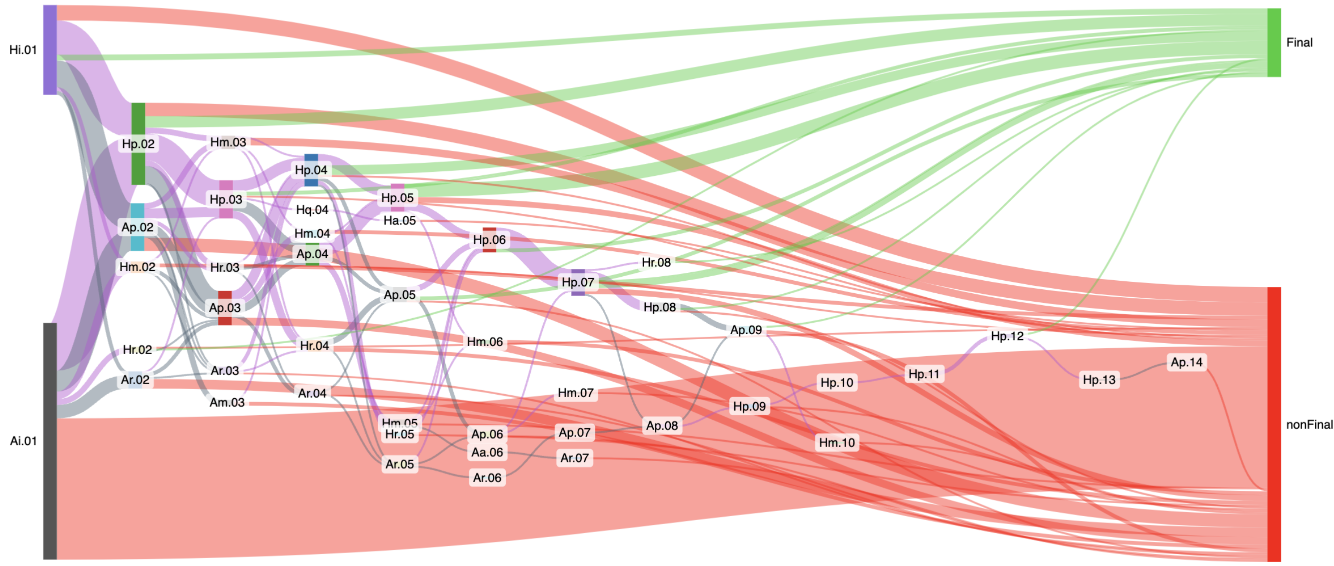

Figure 5: **Summary of Sequences from idea-proposal by Human (*Hi*) or AI (*Ai*), to Outcome (final or nonfinal**. Each event is represented by the codes in Section - e.g., *Ai* is an *idea* proposed by the AI, and *Hp* is a *plus* (supportive) action performed by a human. Orchid colors show actions by humans. Grey colors show actions by AI. Green colors show the last step to human selection as final idea. Red colors show default Outcome of nonfinal. Suffix numbers (e.g., ".04") indicate the step in the Sequence.

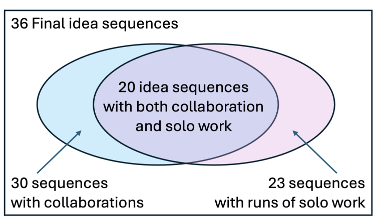

Figure 6: **Venn diagram of final sequences that included at least one instance of human-AI collaboration and at least one instance of a run of two or more repeated actions by the same human or the AI**.

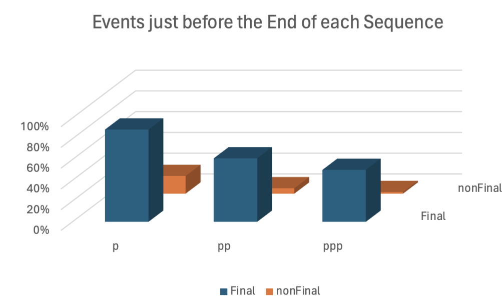

Figure 7: **Runs of *plus* events preceding the end-events of final vs. nonfinal Sequences**.

that 89% final ideas (32 of 36 final ideas) were associated with plus actions, as contrasted with only 33% nonfinal ideas (44 of 135 nonfinal ideas). These are significant differences with $\chi^2_{(1)}$ = 36.480, $p < 0.001$, validated by an approximate Fisher text at < 0.001.

We also examined the events leading up the final action or the nonfinal end of a Sequence. Because the plus action was the most common, we compared action sequences such as *pf*, *ppf*, and *pppf* for final ideas, vs. *p[]*, *pp[]*, and *ppp[]* for nonfinal ideas (where *[]* indicates the end of the nonfinal Sequence). As shown in Figure 7, 89% of final ideas ended with a plus event just before the final event. By contrast, 17% of nonfinal ideas ended with a plus event just before the end of the Sequence. For a run of two plus events, the figures were 61% and 5%, respectively, and for a run of three plus events, the figures were 50% and 1% respectively. These patterns of runs of plus events were significantly different between final and nonfinal sequences, with $\chi^2_{(2)}$ = 7.801, p < 0.03. There were no similar significant results for minus actions or for refine actions.

Becoming Final. Final idea sequences had significantly more plus events, and significantly more and longer sequences of plus events, than nonfinal idea sequences. Supportive (plus) events are partial predictors of an idea being selected as final. Paradoxically, final idea sequences had significantly *fewer* Refine events than nonfinal sequences.

Discussion

Becoming Final

Our Research Question was: How does an idea become final? We analyzed five families of factors that were associated with the final outcome. First, we found that human-proposed ideas had two-to-one advantage over AI-proposed ideas. Second, we saw that ideas that received engagement (actions) from humans or AI were associated with becoming final. Third, we noted that final ideas were more likely to be associated with collaborative actions by two or more participants (including the AI).

Fourth, we also saw that concentrated actions were associated with final outcomes. We speculated that collaboration and solo work might be mutually exclusive. However, we found that just over half of the final ideas were associated with a combination of collaboration and solo work. We observed that supportive (plus) actions were associated with final ideas more than other actions (e.g., minus or refine).

Mixed Initiative Frameworks

Action Vocabularies. In the spirit of evolution of concepts, we compare the actions in our analyses with previously proposed action vocabularies (Deterding et al. 2017; Grabe, Duque, and Zhu 2022; Muller, Weisz, and Geyer 2020; Spoto and Oleynik 2017) as shown in Table 1. Except for Spoto and Oleynik, those vocabularies were based on abstract analyses. Spoto and Oleynik performed their analyses on descriptions of actual systems described in 74 papers. This project allowed us to test some of those abstractions and summaries against the specifics of an evolving set of human practices with an AI participant.

We found utility in only five of the action categories. Our task was relatively simple and transparent, so there was little need for actions of Assemble (or Combine). Our activity was entirely self-contained, obviating the Wait action that generally involved an excursion away from the co-creativity environment (e.g., testing the effectiveness of a proposed solution in a development environment or in the field). We also experienced some ambiguity among some of the action concepts. During a brainstorming session, what is the practical difference between actions of Ideate and Produce? Each idea that is proposed has been produced by a participant. We also realized that there may be little practical difference between actions of Select and Curate. That is, whether participant select or curate seems to depend on the decision rule at the end of the session - select one idea or curate three ideas? An alternate explanation of this ambiguity is that our participants practiced a kind of *serial selection*, without the classification and comparison aspects of curation (Karasti et al. 2016).

Thereby, we believe that the analytic vocabularies should be tested against actual, detailed use cases. Spoto and Oleynik began to do this, but at a high level across many many generative applications. Our more intimate experiences with brainstorming showed us both the limited set of useful actions for our domain, and the potential practical overlaps of some of the concepts.

The Nature of Co-Creativity. Again in the spirit of evolution of concepts, we propose some changes to the 5Cs model of Kantosalo and Takala (2020), in terms of both diversity and sequentiality. We observed a plurality of distinct sequences of actions that humans and AI engaged in (Figures 4- 7). Humans and AI appeared to create these sequences as needed, without a prior plan, and without replication. The sequences are thus not exactly *structures* of a collaboration, because they appear and disappear, and they are subject to mid-sequence variations (Figure 5). We propose to re-use the concept of *ephemerality* from tangible and embodied computing (Döring, Sylvester, and Schmidt 2013), temporary user interfaces (Walsh, Von Itzstein, and Thomas 2014), and temporary data (Michel et al. 2012), to describe these fluidly-changing sequences. We propose a "CIAO" model based on the content and variable dynamics in the sequences:

> A *Conversation* among humans and AI, consisting of their *Ideation* developed through interchangeable, sequential, and ephemerally-structured *Actions* in an *Organizational* context.

In contrast to implications of a collaborative process in the 5Cs model, we propose that the conversation can be more of an optional environment and less of a collective group structure. It suggests a less formal and potentially ephemeral series of spontaneous collaborative *practices*, rather than a structured protocol. We also emphasize the actions that compose these ephemeral sequences, through which known atomic events can be combined by humans and AI into the temporary practices that they need at each moment.

Next Steps

We hope to conduct similar analyses of groups of humans and one or more AI agents, engaged in other co-creative activities. We suspect that only certain actions from the action vocabularies will be relevant in each type of activity. We hope to learn which actions have relatively general applicability, and which have more specialized utility.

We wonder what actions would be important for more structured activities, such as SWOT analyses (Strengths, Weaknesses, Opportunities, and Threats) (Leigh 2009) or SCAMPER analyses (Substitute; Combine; Adapt; Modify, minify or magnify; Put to another use; Eliminate or elaborate; and Reverse) (Hassan 2023). Perhaps our finding of ephemerality will require revision for activities with more prescriptive work practices.

Contributions

We conducted one of the first experiments with multiple humans and a LLM-based AI, collectively engaged in the co-creative task of online brainstorming. We demonstrated that humans and AI can work collaboratively to propose and improved shared ideas in a brainstorming session. Through a detailed analysis of who or what acted, and the nature of specific actions, we identified five families of factors that were associated with successful ideas in the brainstorm. Using these patterns in the data, we proposed evolutionary developments in mixed initiative creative frameworks, including action vocabularies and a CIAO model of human-AI interactive co-creativity. We anticipate further developments by ourselves and others with LLM-based AI agents collaborating in practical human workplace co-creative activities.

References

Bouschery, S. G.; Blazevic, V.; Roberts, D.; and Piller, F. T. 2022. Artificial intelligence as an actor in innovation teams: An assessment of the gpt-3 language model. In *Academy of Management Proceedings*, volume 2022, 11505. Academy of Management Briarcliff Manor, NY 10510.

Bouschery, S. G.; Blazevic, V.; and Piller, F. T. 2023. Ai-augmented creativity: Overcoming the productivity loss in brainstorming groups. In *Academy of Management Proceedings*, volume 2023, 11938. Academy of Management Briarcliff Manor, NY 10510.

Chiou, L.-Y.; Hung, P.-K.; Liang, R.-H.; and Wang, C.-T. 2023. Designing with ai: An exploration of co-ideation with image generators. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, 1941–1954.

Deterding, S.; Hook, J.; Fiebrink, R.; Gillies, M.; Gow, J.; Akten, M.; Smith, G.; Liapis, A.; and Compton, K. 2017. Mixed-initiative creative interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 628–635.

Döring, T.; Sylvester, A.; and Schmidt, A. 2013. A design space for ephemeral user interfaces. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, 75–82.

Dugosh, K. L.; Paulus, P. B.; Roland, E. J.; and Yang, H.-C. 2000. Cognitive stimulation in brainstorming. *Journal of personality and social psychology* 79(5):722.

Farrell, R.; Lenchner, J.; Kephart, J.; Webb, A.; Muller, M.; Erickson, T.; Melville, D.; Bellamy, R.; Gruen, D.; Connell, J.; et al. 2016. Symbiotic cognitive computing. *ai Magazine* 37(3):81–93.

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

Franceys, R.; Guinaldo, T.; Leitner, C.; Nyangoka, O.; Thomas, V.; Zeilinger, J.; and Hutchings, P. 2017. Financial flow diagrams to promote policy-making, based on 20 community management case studies from india. https://repository.lboro.ac.uk/articles/conference_contribution/Financial_flow_diagrams_to_promote_policy-making_based_on_20_community_management_case_studies_from_India/9589427/files/17228549.pdf.

Geerts, J.; de Wit, J.; and de Rooij, A. 2021. Brainstorming with a social robot facilitator: Better than human facilitation due to reduced evaluation apprehension? *Frontiers in Robotics and AI* 8:657291.

Glăveanu, V. P. 2014. *Distributed Creativity: What Is It?* Cham: Springer International Publishing. 1–13.

Grabe, I.; Duque, M. G.; and Zhu, J. 2022. Towards a framework for human-ai interaction patterns in co-creative gan applications. In *Proceeding of the 3rd Workshop on Human-AI Co-Creation with Generative Models (HAI-GEN '22) at ACM IUI Workshops*.

Hassan, S. 2023. Scamper as a creative idea generation method: Case study on graphic design students. *Inf. Sci. Lett* 12:1417–1428.

Hernandez, A. G.; Lupton, R. C.; Williams, C.; and Cullen, J. M. 2018. Control data, sankey diagrams, and exergy: Assessing the resource efficiency of industrial plants. *Applied Energy* 218:232–245.

Hollan, J.; Hutchins, E.; and Kirsh, D. 2000. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7(2):174–196.

Horvitz, E. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 159–166.

John-Steiner, V. 2000. *Creative collaboration*. Oxford University Press.

Kantosalo, A., and Takala, T. 2020. Five c's for human-computer co-creativity-an update on classical creativity perspectives. In *ICCC*, 17–24.

Karasti, H.; Millerand, F.; Hine, C. M.; and Bowker, G. C. 2016. Knowledge infrastructures: part i. *Science & Technology Studies* 29(1):2–12.

Koch, J.; Lucero, A.; Hegemann, L.; and Oulasvirta, A. 2019. May ai? design ideation with cooperative contextual bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Kuss, P., and Leenes, R. 2020. The ghost in the machine-emotionally intelligent conversational agents and the failure to regulate'deception by design'. *SCRIPTed* 17:320.

Laitinen, K.; Laaksonen, S.-M.; and Koivula, M. 2021. Slacking with the bot: Programmable social bot in virtual team interaction. *Journal of Computer-Mediated Communication* 26(6):343–361.

Lamer, A.; Laurent, G.; Pelayo, S.; Amrani, E.; Chazard, E.; and Marcilly, R. 2020. Exploring patient path through sankey diagram: a proof of concept. *Studies in health technology and informatics* 270.

Lavrič, F., and Škraba, A. 2023. Brainstorming will never be the same again—a human group supported by artificial intelligence. *Machine Learning and Knowledge Extraction* 5(4):1282–1301.

Leigh, D. 2009. Swot analysis. *Handbook of Improving Performance in the Workplace: Volumes 1-3* 115–140.

Liikkanen, L. A.; Kuikkaniemi, K.; Lievonen, P.; and Ojala, P. 2011. Next step in electronic brainstorming: collaborative creativity with the web. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM. 2029–2034.

Maher, M. L.; Paulini, M.; and Murty, P. 2011. Scaling up: From individual design to collaborative design to collective design. In *Design Computing and Cognition'10*, 581–599. Springer.

Maher, M. L. 2012. Computational and collective creativity: Who's being creative? In *ICCC*, 67–71.

Michel, J.; Julien, C.; Payton, J.; and Roman, G.-C. 2012. A spatiotemporal model for ephemeral data in pervasive computing networks. In *2012 IEEE International Conference* *on Pervasive Computing and Communications Workshops*, 179–184. IEEE.

Millwood, A., and Dias-Taguatinga, C.-C. 2023. Ai image generation tools as an aid in brainstorming architectural visual designs, thesis, stockholm university.

Mongeau, P. A., and Morr, M. C. 1999. Reconsidering brainstorming. *Group Facilitation* 1:14.

Muller, M., and Weisz, J. 2023. Analogies-based design using a generative ai application: A play in three acts. http://studiolab.ide.tudelft.nl/studiolab/genai-dis2023/files/2023/07/DIS_2023_workshop___analogies-2.pdf.

Muller, M.; Candello, H.; and Weisz, J. 2023. Interactional co-creativity of human and ai in analogy-based design. In *International Conference on Computational Creativity*.

Muller, M.; Weisz, J. D.; and Geyer, W. 2020. Mixed initiative generative ai interfaces: An analytic framework for generative ai applications. In *Proceedings of the Workshop The Future of Co-Creative Systems-A Workshop on Human-Computer Co-Creativity of the 11th International Conference on Computational Creativity (ICCC 2020)*.

Muller, M. 2023. Exploring human-ai co-creativity under human control: Framing, reframing, brainstorming, and future challenges (workshop keynote). *CEUR Workshop Proceedings 3547*.

Nijstad, B. A.; Stroebe, W.; and Lodewijkx, H. F. 2003. Production blocking and idea generation: Does blocking interfere with cognitive processes? *Journal of experimental social psychology* 39(6):531–548.

Oh, C.; Song, J.; Choi, J.; Kim, S.; Lee, S.; and Suh, B. 2018. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.

Orit Shaer, Angelora Cooper, A. K., and Mokryn, O. 2024. Toward enhancing ideation through collaborative group-ai brainwriting. In Soto, A., and Zangerle, E., eds., *Joint Proceedings of the ACM IUI 2024 Workshops co-located with the 29th Annual ACM Conference on Intelligent User Interfaces (IUI 2024), Greenville, South Carolina, USA, March 18, 2024*, volume 3660 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Paulus, P. B., and Brown, V. R. 2003. Enhancing ideational creativity in groups. *Group creativity: Innovation through collaboration* 110–136.

Paulus, P. B., and Dzindolet, M. 2008. Social influence, creativity and innovation. *Social Influence* 3(4):228–247.

Paulus, P. B.; Dzindolet, M.; and Kohn, N. W. 2012. Collaborative creativity—group creativity and team innovation. In *Handbook of organizational creativity*. Elsevier. 327–357.

Paulus, P. B. 2015. Electronic brainstorming research and its implications for e-planning. *International Journal of E-Planning Research (IJEPR)* 4(1):42–53.

Rezwana, J., and Maher, M. L. 2021. Cofi: A framework for modeling interaction in human-ai co-creative systems. In *ICCC*, 444–448.

Salles, A.; Evers, K.; and Farisco, M. 2020. Anthropomorphism in ai. *AJOB neuroscience* 11(2):88–95.

Santanen, E. L.; Briggs, R. O.; and De Vreede, G.-J. 2000. The cognitive network model of creativity: A new causal model of creativity and a new brainstorming technique. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 10–pp. IEEE.

Seering, J.; Luria, M.; Kaufman, G.; and Hammer, J. 2019. Beyond dyadic interactions: Considering chatbots as community members. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Shaer, O.; Cooper, A.; Mokryn, O.; Kun, A. L.; and Shoshan, H. B. 2024. Ai-augmented brainwriting: Investigating the use of llms in group ideation.

Shneiderman, B., and Muller, M. 2023. On ai anthropomorphism, *Medium* 2023-04-10. https://medium.com/human-centered-ai/on-ai-anthropomorphism-abff4cecc5ae.

Spoto, A., and Oleynik, N. 2017. Library of mixed-initiative creative interfaces. http://mici.codingconduct.cc/.

Stroebe, W.; Nijstad, B. A.; and Rietzschel, E. F. 2010. Beyond productivity loss in brainstorming groups: The evolution of a question. In *Advances in experimental social psychology*, volume 43. Elsevier. 157–203.

Suh, M.; Youngblom, E.; Terry, M.; and Cai, C. J. 2021. Ai as social glue: uncovering the roles of deep generative ai during social music composition. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, 1–11.

Walsh, J. A.; Von Itzstein, S.; and Thomas, B. H. 2014. Ephemeral interaction using everyday objects. In *Proceedings of the Fifteenth Australasian User Interface Conference-Volume 150*, 29–37.

Wieland, B.; de Wit, J.; and de Rooij, A. 2022. Electronic brainstorming with a chatbot partner: A good idea due to increased productivity and idea diversity. *Frontiers in Artificial Intelligence* 5:880673.

Yu-Han, C., and Chun-Ching, C. 2023. Investigating the impact of generative artificial intelligence on brainstorming: A preliminary study. In *2023 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*, 193–194. IEEE.