Toward Enhancing Ideation through Collaborative Group-AI Brainwriting

Orit Shaer1,*,†, Angelora Cooper1, Andrew L. Kun3 and Osnat Mokryn4

1Wellesley College, 106 Central st., Wellesley, 02481, USA
3University of New Hampshire, Durham, NH 03824, USA
4University of Haifa, Haifa, Israel

Abstract
This paper introduces a collaborative group-AI framework for enhancing ideation through co-creation. The proposed framework integrates LLMs into the creative process to support both the divergence stage of idea generation and the convergence stage of evaluation and selection of a few chosen ideas. We describe the framework and the tools we designed to implement it as well as summarize findings from its evaluation with novice designers - students in an advanced interaction design course. Our findings suggest that the framework could enhance both the ideation process and its outcome through human-AI co-creation.

Keywords
LLM, Brainwriting, Group ideation, Human-AI collaboration, CEUR-WS

1. Introduction

The growing availability of generative AI technologies including large language models (LLMs) and image models [1] have profound impact on the work of designers and other creative professionals [2, 3, 4]. Creative collaborative workflows often follow two phases. In an initial divergence phase, teams generate a broad range of possible ideas. In a following convergence phase, all generated ideas are reviewed and evaluated by the team members, with the goal of identifying and choosing the few ideas that the team will pursue further. We are interested in investigating how LLMs can be integrated effectively into both the divergence and convergence ideation phases to enhance teams’ creativity. We expect that in the divergence phase, LLMs can be used to improve ideas generated by people, and to suggest new ideas. In the convergence phase, LLMs can help determine which ideas are more relevant, innovative and insightful, as well as to aid in the further development of the chosen ideas.

To explore this question we devised a collaborative group-AI ideation framework, which incorporated an LLM as an enhancement into a group’s creative process. The LLM does not replace human input but rather adds to it and augments it. The proposed group-AI framework
draws upon the brainwriting process [5], which is an alternative or complementary method to the widely used face-to-face group brainstorming process.

During a successful brainstorming session, participants generate new ideas by drawing on each other’s suggestions [6]. However, despite the prevalence of group brainstorming, it is shown that a greater number of ideas and better quality ideas are generated when individuals brainstorm independently [7]. When individuals work alone they tend to consider many different potential solutions, but when team members work together they often consider fewer alternative solutions because of factors such as peer judgment, free riding, and production blocking [8].

Brainwriting aims to address these shortcomings through a parallel (rather than a sequential) process [9] - all participants write down their ideas in response to a given prompt in parallel, before sharing their ideas with others. Only after all participants wrote their ideas, participants review others’ ideas and then add new ones by either individually writing additional ideas or through discussion and collaboration. It is found that the number of high quality ideas generated from brainwriting sessions often exceeds face-to-face brainstorming [10].

Our group-AI ideation framework is shown in Figure 1. We introduced an earlier version of this framework and initial evaluation in [11]. The earlier version did not include the custom interfaces and the thinking engines introduced here. It draws upon Paulus and Yang’s [10] suggestion of a two-phase process for ideation. In the divergence stage of our multi-step process, group members first generate their own ideas and add them to a shared online whiteboard. Then, group members review and interact with their collective ideas while prompting an LLM for new ideas that will enhance their initial set of ideas. In the convergence stage, the LLM is used to assist group members to evaluate their ideas and to narrow the list of ideas to a few selected chosen ideas. Finally, group members use the LLM to assist in further developing the selected ideas.
In the following, we describe our collaborative group-AI brainwriting framework in detail and describe how we have evaluated it with novice designers.

2. Related Work

There is an emerging research exploring how co-creation with generative AI could support and enhance interaction design [4] and what co-creation practices might look like for ideation [12, 13, 14], prototyping, making, and programming [15, 16, 17]. Such human-AI co-creation processes should also be considered within the context of emerging theories about posthumanism, post-human, and more-than-human interaction design [18, 19, 20, 21], which highlight and explore possibilities to distribute agency in design between human and non-human agents.

Within the domain of ideation, researchers showed that collaborative approaches could lead to more creative solutions by exposing people to different perspectives and exploring new connections through diverse ideas [22, 23, 24, 25]. Several online platforms for large-scale ideation, were designed to leverage diversity of ideas by implementing methods to select and present creative and diverse ideas [25]. Our focus, rather than supporting large-scale ideation, is to enhance small groups (3-4 people) ideation process through the use of LLMs.

Online visual workspaces such as Miro [26], ConceptBoard [27] and Mural [28] offer support and template for both remote and co-located ideation processes, and have integrated LLMs functionality as part of their products. However, additional research is needed to identify the merits and limitations of integrating LLMs into the ideation processes. Shin et al [29] led a CHI 2023 workshop to explore the integration of AI in human-human collaborative ideation. Our goal is to add to the body of knowledge on collaborative group-AI ideation.

3. Collaborative Group-AI Brainwriting Framework Design

We describe here how we integrate the LLM into the ideation phases.

3.1. Brainwriting Divergence Stage

The goal of the divergence stage is for participants to produce a wide range of different ideas [30]. The quantity of ideas generated in this stage is important because people are more likely to find quality ideas when selecting from a large number of ideas [31]. Our approach is to enhance this stage by treating the LLM as an additional team member - contributing additional ideas rather than replacing humans in the idea generation process.

3.1.1. Brainwriting using an online whiteboard

In this modified Brainwriting process [5] group members sit together around a shared table, but write their ideas separately, in parallel, on an online whiteboard (we used Conceptboard [27]). Each participant selects a color on the board, then the group sets a timer for 3 minutes and use that time to write ideas independently. Each group member writes at least three ideas relevant to the problem statement and place them on the board using colored coded sticky notes. Then participants are asked to repeat this process until each group member wrote at least six
ideas. Figure 2 shows the modified Conceptboard template we used for the Brainwriting activity, populated with ideas generated by one of the student teams in our study. The Conceptboard template we use is based on the Conceptboard’s remote Brainwriting template [32].

3.1.2. Enhancing ideas with an LLM

In this step, the group uses an LLM-powered tool to generate additional ideas. The LLM plays the role of an additional team member. The generated ideas are added into sticky notes on the board. We modified the original Brainwriting template offered by Conceptboard to reflect this new framework for Brainwriting with LLM (see Figure 2).

The group reviews all initial ideas (human and LLM generated), discusses them, and develops together, with the help of the LLM, new ideas that add to or build upon the existing preliminary ideas. These ideas are added to an area on the board dedicated to collaborative ideas.
3.2. Brainwriting Convergence Stage

The goal of the convergence phase is to evaluate and select a small set of quality ideas and proceed to develop the details of the chosen ideas incrementally, towards a solution [30].

Involving LLMs in idea selection stage holds the promise of increased speed of idea evaluation, as well as the opportunity for the AI to support the creative efforts of humans by providing feedback (e.g [33]), and in this work we explore using an LLM to evaluate the written ideas generated by teams comprising of humans and another LLM. Domonik [34] shows that AI evaluation could also improve human ideation by reducing evaluation apprehension - the situation where a human will withhold an idea for fear of being evaluated negatively.

3.2.1. LLM Powered Evaluation

As group members review, discuss and evaluate the proposed ideas, they consult with an LLM-powered evaluation engine. The evaluation engine aims to provide an additional perspective rather than automating the idea selection process. We developed a GPT-4 powered evaluation engine, which builds on the approach of Dean et al. [35] for evaluating the quality of ideas. The evaluation engine uses the dimensions of novelty (which we call innovation) and relevance to evaluate ideas. In addition, we chose a third criterion, insightfulness, based on Dyer et al.’s research on the origin of innovative ventures [36]. Figure 3 shows a prototype of the GPT-powered evaluation engine. The evaluation engine provides users with a numerical score (on a Likert scale of 5 for each of the three dimensions): innovation, relevance, and insightfulness. For each score there is a qualitative feedback, which explains the score.
### 3.2.2. Developing and Refining Ideas with LLM

Once a small set of ideas is selected, group members use an LLM-powered thinking engine, which is designed to assist in considering different aspects of each idea. Tversky and Chou suggest that shifting attention between different problems enhances creativity [30]. Our thinking engine adopts an approach similar to "Six Thinking Hats" [37], where different prompts are constructed in the back end, each defining a different persona for the LLM and hence leading to considering different aspects and to representing different perspectives. Figure 4 shows a prototype of our GPT-4 powered thinking engine.

![Figure 4: The design prototype of the GPT-4-powered "Six Hats" thinking engine.](image-url)

### 4. Framework Evaluation

We have been evaluating the collaborative Group-AI design framework by using it in advanced undergraduate-level interaction design courses.

First, we ran a user study that uses the framework in the divergence stage for idea generation. We evaluated the process itself, and its outcomes - the set of ideas.

Second, we studied the potential of using the GPT-4 evaluation engine for assisting in evaluating and selecting ideas.

This evaluation process and our findings are described in detail in [11]. Here, we provide a summary of our findings.

We are currently in the process of evaluating the complete collaborative Group-AI Brain-writing framework and custom interfaces with novice designers by deploying it in advanced
interaction design courses.

4.1. Evaluation of the Divergence stage - the Collaborative Brainwriting session

In Spring 2023, we conducted a 70-minutes Brainwriting session with 16 college students (0 men, ages 18-23) who were enrolled in an advanced undergraduate course on tangible interaction design. Considering the challenges interaction designers face when working with AI as a design material [38, 39, 40, 41, 42], this course aims to integrate co-creation and critical engagement with generative AI into its learning goals. The course’s learning goals and approach to co-creation with generative AI is described in [4].

The students were divided into 5 project teams of 3-4 students each. The goal for the session was for students to develop project ideas for a semester-long group project. The brief for the project was: “design a novel tangible user interface, which helps support the productivity, creativity, and well-being of people who work or study in mobile environments.” Table 1 shows the number of ideas generated by each team. The students submitted a link to their Conceptboard used for the Brainwriting, as well as all their GPT-3 prompts used for idea generation. At that time we selected to use GPT-3 because it was freely available and accessible to all students. At the end of the session, the students were asked to rate the ideas: their own, GPT-3 generated, and the collaborative ideas, as a means to narrow down the idea pool and engage in a selection process. The ideas were rated on a Likert scale along the three chosen evaluation criteria of relevance, innovation and insightfulness. After the session, each team chose an idea for their semester-long project. Finally, we asked students about their experience Brainwriting with GPT-3 both immediately after the session, as well as again at the end of the semester.

4.1.1. Summary of Findings

In their responses, after the brainwriting session, 50% of students perceived GPT-3 as helpful because it provided a unique or expanded perspective on the problem statement and its possible solutions. 44% shared that it significantly assisted them in generating new ideas. At the end of the semester, 50% of the students mentioned that GPT-3 contributed to reshaping and enhancing their project by elaborating on their concepts, proposing new characteristics, and tackling particular challenges. 31% of students pointed out that GPT-3 tends to be redundant and lacked creativity.

The ideas selected by each group for their final project were mostly created by combining an idea generated by team members and an idea suggested or enhanced by the LLM.

Semantic clustering analysis of Human- and GPT-3-Generated ideas indicated that humans tended to allude to abstract concepts and refer to objects in a general way, while the ideas generated by GPT-3 were more concrete and included material and technical details. For example, the term “device” appears almost exclusively in GPT-3-Generated ideas, which often also reference their “users”. In Human-Generated ideas, the reference is to “people”, and the term “wearable” appears only in human ideas.

The prompt analysis reveals that students combined approaches when interacting with GPT-3, typically starting with a broad request for ideas, then requesting solutions for a concrete
problem, or asking for additional details regarding the usage, features, and/or capabilities of a specific idea. These results explain, to some extent, the higher level of details we found in GPT-3-Generated ideas.

Table 1
The number of ideas created per team: Human-Generated, GPT-3-Generated, Collaboratively-Generated, and total.

<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>GPT-3</th>
<th>Collaborative</th>
<th>Total # of ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team 1</td>
<td>20</td>
<td>4</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>Team 2</td>
<td>18</td>
<td>11</td>
<td>11</td>
<td>40</td>
</tr>
<tr>
<td>Team 3</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>Team 4</td>
<td>24</td>
<td>6</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>Team 5</td>
<td>18</td>
<td>6</td>
<td>3</td>
<td>27</td>
</tr>
</tbody>
</table>

4.2. Assessing the Feasibility of an LLM-based Evaluation Engine

We assessed the feasibility of using an LLM to assist in idea evaluation in the convergence phase separately from the user study. Our evaluation was conducted after the student deadline for choosing their final ideas. To evaluate whether an LLM can help in the convergence phase, where ideas are evaluated and a few are selected, we assessed (a) whether LLMs' evaluations are consistent, and (b) how they compare with evaluations made by expert reviewers (HCI researchers and faculty) and by novice reviewers (peer).

All ideas created during the Brainwriting process: Human-Generated, GPT-3-Generated, and Collaboratively-Generated, were evaluated by 3 Experts, 6 Novices, and the GPT-4 evaluation engine. All evaluations used the same 1 to 5 Likert Scale for Relevance, Innovation, and Insightfulness. Both Novice and Expert reviewers were given the same criteria definition and scale value anchors given to the GPT-4 evaluation engine. The ideas given to the reviewers were arranged in a random order and there was no identifying information regarding the source of the idea (human or GPT-3). The GPT-4 engine was prompted to repeat each evaluation 30 times (29 rounds were completed successfully), each evaluation was conducted in a new context.

4.2.1. Summary of Findings

To assess the internal consistency of the 29 GPT-4 evaluations for the ideas on the three criteria of Relevance, Innovation, and insightfulness we treated the evaluations as questionnaire items and analyze them with Fleiss’ Kappa coefficients to evaluate rater agreement. Our analysis shows a moderate level of consistency in GPT-4’s performance (all Fleiss’ Kappa values surpassing the 0.4 threshold) across the three criteria.

To quantify the relationship between the rankings provided by Experts, Novices, and GPT-4, we computed Pearson correlation coefficients. The comparison indicated a moderate positive linear relationship among the three rater groups. Thus, showing that GPT-4’s ranking of ideas is generally in agreement with the Experts’ and Novices’ rankings.

Finally, we observed that the GPT-4 evaluation engine gave high ratings to all of the ideas that were ultimately chosen for a project by student teams. The fact that none of the chosen
ideas received low ratings by GPT-4 is encouraging - it means that, if GPT-4 had been used to provide feedback for teams during the ideation process, it would not have filtered out ideas that were considered to be good by the teams.

4.3. Evaluating "In the Wild"

Based on the generally positive results from our evaluation with students of the Human-AI brainwriting method and our feasibility assessment of a GPT-4 evaluation engine, we are currently iterating on the design and development of custom GPT-4 powered interfaces for the convergence and divergence stages. These tools (see prototypes in Figures 3 and 4), implement back-end prompt engineering. This approach could potentially address the challenges reported by students with designing effective prompts and lead to more effective co-creation processes.

We plan to evaluate the use of these tools “in the wild” by deploying them in advanced interaction design courses across several institutions. In the long term, in addition to co-located participants, we also plan to explore how remote participants can utilize LLMs in the creative process (cf. [43]). Our aim is to evaluate the both the co-creation process itself as well as its products.

5. Conclusion

We expect that human-AI co-creation processes will reshape creative work in the near future. In this work we explore one potential scenario of such a collaboration, using LLMs for enhancing group ideation. Our focus is on Brainwriting - a framework for ideation, where we explore how LLMs can enhance the ideas generated by a creative team. Rather than replacing team members, we view the role of the AI as an additional team member capable of providing additional perspectives and details.

Our results so far indicate that LLMs can be useful for supporting both the divergence and convergence stages of the process of generating ideas. The educational settings in which we conducted our evaluation also shows that the collaborative group-AI brainwriting framework we propose could serve as a tool for both educators and novice designers [? ].

Acknowledgments

References


[34] D. Siemon, Let the computer evaluate your idea: evaluation apprehension in human-computer collaboration, Behaviour & Information Technology 42 (2023) 459–477.


[42] H. Subramonyam, C. Seifert, E. Adar, Towards a process model for co-creating ai experi-
[43] A. A. Ansah, Y. Xing, A. V. Kamaraj, D. Tosca, L. Boyle, S. Iqbal, A. L. Kun, J. D. Lee, M. Pahud, O. Shaer, "I need to respond to this"–contributions to group creativity in remote meetings with distractions, in: 2022 Symposium on Human-Computer Interaction for Work, 2022, pp. 1–12.