Collaborative Canvas: A Tool for Exploring LLM Use in Group Ideation Tasks

Gabriel Enrique Gonzalez[1], Dario Andres Silva Moran[1], Stephanie Houde[2], Jessica He[2], Steven I. Ross[2], Michael Muller[2], Siya Kunde[2] and Justin D. Weisz[2]

[1]*IBM Argentina, AR*

[2]*IBM Research AI, USA*

Abstract

We present the *Collaborative Canvas*, a prototype tool for exploring ways for groups to interact with large language models (LLMs) in ideation tasks. *Collaborative Canvas* provides a shared, graphical canvas in which multiple parties – human and LLM – can share ideas in the form of virtual "sticky notes" that can be moved around the canvas. The development of *Collaborative Canvas* raised numerous issues about the role of an LLM in group interactions: is it useful, what role does it play within the group's workflow, and how do people interact with generated content? A preliminary examination of the *Collaborative Canvas* shows that users found the generative capabilities to be useful, although they preferred to review and filter generated content before sharing it with the group. Users also speculated that the role of the AI could extend into facilitating group brainstorming rather than being confined to idea generation. Our work motivates the study of human-AI co-creation in group settings beyond dyadic interactions.

Keywords

Generative AI, Group Brainstorming, Group Ideation, Large Language Models, Co-creativity

1. Introduction

Researchers have recently begun exploring the role of generative AI for various tasks within design, including persona generation [1], ideation [2], early prototyping [3], and participatory design [4]. Popular shared canvas tools such as Mural[1] and Miro[2] have also begun introducing generative AI features within their applications. These features enable users to leverage AI to generate ideas, create diagrams and presentations, answer questions, summarize content, and cluster sticky notes into themes.

Many studies of human-AI co-creativity and human use of generative AI focuses on the interactions between a *single* human user with a generative AI application. However, given the *collaborative* and *real-time* nature of shared canvas tools, new research questions arise around

[1]Mural. https://mural.co

[2]Miro. https://miro.com

the role of generative AI in group work: do people in a group find it useful to have an integrated generative "assistant," what role does it play within the group's workflow, and how do people interact with generated content?

We developed the *Collaborative Canvas* as a way to prototype different forms of interaction with generative AI, and more specifically, with large language models. In developing our prototype, we uncovered important design considerations regarding the ownership and visibility of generated content and the UX mechanisms used to invoke the generative functionality. Our work seeks to identify new research avenues for studying the impact that generative AI has in group settings, pushing research in human-AI co-creativity beyond the examination of dyadic interactions.

2. Related Work

Wang et al. [5] recently described the challenges of human-AI collaboration by reminding the HCI community that, "Interaction is not the same as Collaboration" ([5, p.1]. Other researchers, including Aragon et al. [6], Inkpen et al. [7], and Shneiderman [8] have made similar arguments. The concept of AI agents acting as "teammates" has recently gained interest (e.g., [9]), but research in human-AI co-creation often focuses on the "team" as being composed of one human and one AI agent (e.g. [10, 11, 12, 13, 14, 15, 16]).

Despite the prevailing focus on human-AI dyads, some researchers have begun exploring the use of generative AI in larger team settings. Gan et al. [17] introduced an AI-Mediated Group Ideation tool that acts as a creative mediator in an architectural ideation process. In their study, they found that the AI mediator did have utility, but they also identified challenges in integrating it within the workflow in ways that were not distracting. Cvetkovic et al. [18] derived a number of design principles for a conversational assistant that collects and summarizes ideas from a group ideation session. Their study also found utility in having the assistant participate in group ideation, but identified challenges with knowing when it should make a contribution and how much to contribute. Lavrič and Škraba [19] examined the brainstorming capabilities of OpenAI's GPT-3.5-turbo model and found that the ideas it produced were useful. Although their examination was of ideas that were solely generated by the LLM, the authors concluded, "the hybrid process of generating ideas should be tested by real committees, where the OpenAI GPT API would be applied in combination with human agents. This might contribute to a better set of innovative ideas..." [19, p.1297].

3. Collaborative Canvas

We designed and implemented *Collaborative Canvas* to explore ideas around group collaboration with generative AI. Motivated by the emerging use of generative AI within the design profession, we desired an environment that would be useful to designers for conducting various activities within the practice of design thinking. As the designers within our own organization make heavy use of Mural, we initially explored whether we could incorporate generative capabilities using its API[3]. However, we found that we were unable to prototype many types of interactions

---

[3]Our explorations began before the introduction of Mural AI: https://www.mural.co/blog/announcing-mural-ai

with Mural's API[4] and pivoted to building our own tool.

We show a screenshot of the *Collaborative Canvas* in Figure 1. As we aimed not to replicate a commercial product, it possesses a minimally-useful set of functionality commonly found in shared canvas tools:

- Multiple users are able to join the shared canvas environment and interact with it in real time. A list of connected users shows who is currently in the environment.
- Users are able to add, select, and remove sticky notes.
- Users are able to change the size, color, and position of sticky notes.
- Users are able to zoom in and out within the canvas.
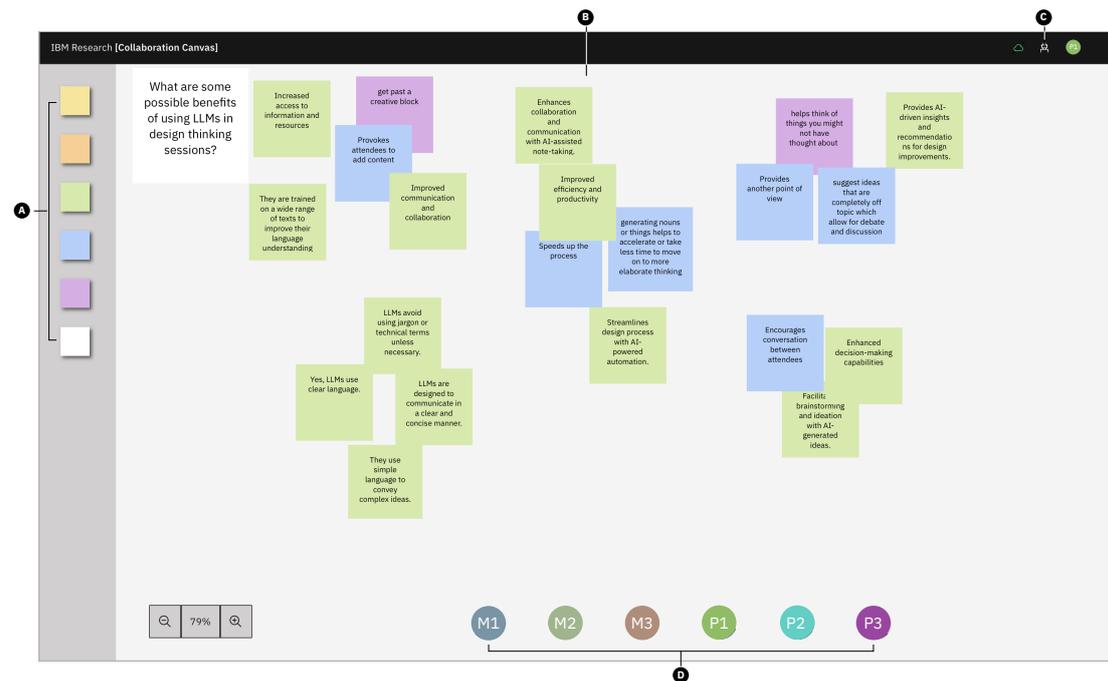
Figure 1: Collaborative Canvas user interface. The shared canvas provides a space for multiple users to add, remove, and move around virtual sticky notes. (A) A sidebar provides users with the ability to add new sticky notes in various colors. (B) Sticky notes can be placed and moved around the canvas. (C) A separate panel can be revealed to show a private scratchpad for generating new sticky notes with an LLM off of the main canvas (shown in Figure 3). (D) User profile images represent the users who are currently connected to the canvas (these images have been anonymized for this paper).

In addition to providing core shared canvas functionality, *Collaborative Canvas* also offers several different ways to invoke a large language model to generate new content:

---

[4]At the time, Mural's API did not support their embedded chat feature, eliminating the possibility of examining conversational interactions. In addition, the API did not provide support for reading the selection state of a user, eliminating the possibility of selecting a group of sticky notes and invoking a generative model to, for example, summarize their contents.

- New sticky notes can be generated and added to the main canvas from a textual prompt in a modal (Figure 2). These notes are visible to all users when they are generated.
- New sticky notes can be generated within a private scratchpad (Figure 3A & B). Users are then able to drag sticky notes from the scratchpad into the shared canvas space (Figure 3C).
- Users can progressively refine the content of generated sticky notes through follow-up requests. The sequence of requests are preserved so that notes can be regenerated in the context of previous requests (Figure 3D).
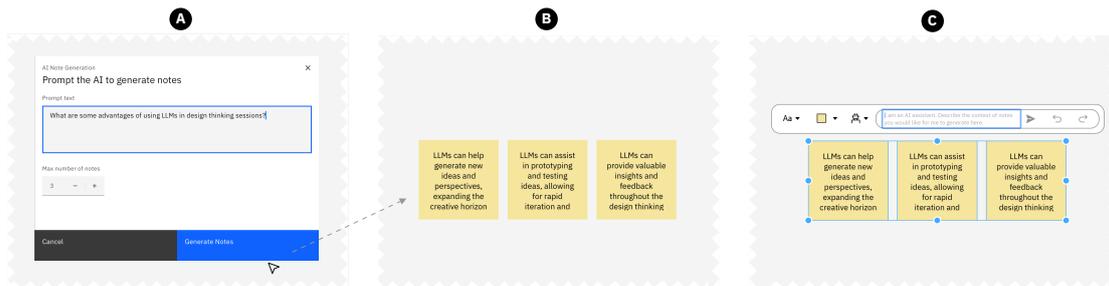

Figure 2: Generative AI capabilities within the shared canvas. Users can generate sticky notes directly in the shared canvas view. (A) By right-clicking on the canvas and choosing "Generate" from the contextual menu, a user can access a dialog for generating sticky notes. The dialog allows the user to specify their prompt and request that a specific number of sticky notes are generated. (B) After clicking the "Generate Notes" button, new sticky notes are created (with an interstitial animation to hide LLM inference latency) in the main canvas. (C) When selecting one or more sticky notes, a pop-up toolbar allows users to re-generate the selected sticky notes with a new prompt. In this way, users are able to make refinements to their own (or generated) sticky notes, such as "provide a condensed summary" or "come up with new ideas."

3.1. Implementation

We implemented the *Collaborative Canvas* using the Svelte[5] framework for the front-end UI and Python with FastAPI[6] for the back end. The back end was responsible for communicating with an internal version of the IBM watsonx.ai platform where the LLM was hosted. While this platform provides API access to a number of state-of-the-art LLMs, we chose the Llama 2 [20] model for our prototyping and experiments. We developed a prompt for this model to set the context for users' actions, including generating sticky notes and progressively refining generated responses. This prompt also instructed the model to format its responses in a way that was easy to parse into separable chucks by the front end (e.g. to create multiple sticky notes from a single LLM response). We list the LLM prompt used by *Collaborative Canvas* in Appendix A. In Figure 4, we illustrate the flow that occurs from the moment a user requests the LLM to generate a group of sticky notes until they are generated.

---

[5]Svelte. https://svelte.dev
[6]FastAPI. https://fastapi.tiangolo.com
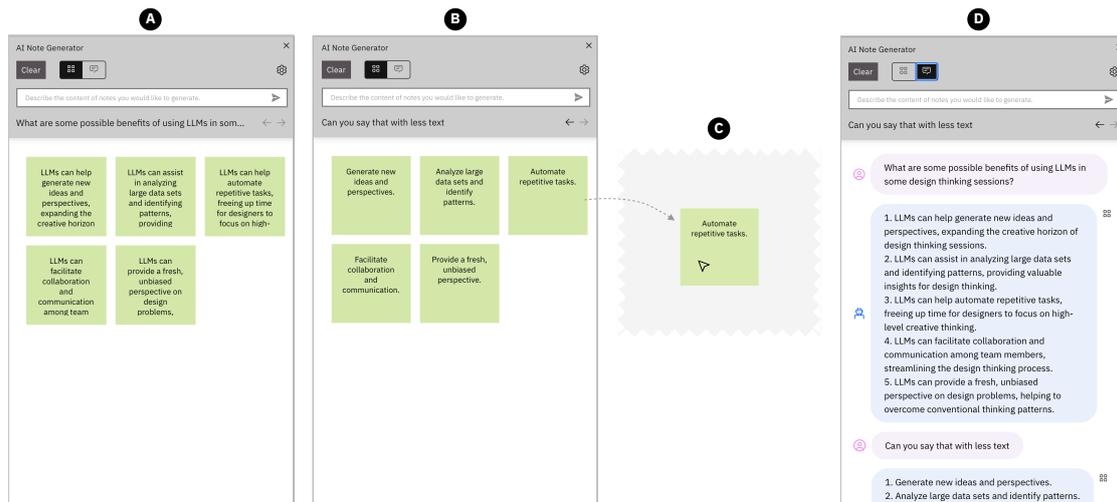
Figure 3: Personal scratchpad. As an alternative to generating public sticky notes on the shared canvas, we also developed a personal scratchpad for generating sticky notes, enabling users to review their content before sharing them with the group. (A) The scratchpad provides a separate space to prompt the LLM and generate new sticky notes. (B) The current state of the scratchpad is preserved between subsequent prompts, so a user can request that the sticky notes that were previously generated "say that with less text." (C) Users can choose which sticky notes in their scratchpad they would like to share with the group by dragging them from the scratchpad to the main canvas. (D) Users can view their progressive prompts and regenerated responses in a chat-like sequential view that can be clicked to navigate back to older versions of generated note sets.

4. Preliminary evaluation

We conducted a preliminary evaluation of the *Collaborative Canvas* to understand the extent to which its different generative capabilities could be used within a group brainstorming session. We recruited 8 design professionals within our organization who facilitate design thinking workshops as part of their work. Our goal was to learn how these domain experts, who currently make extensive use of shared canvas tools for collaborative work, might use the integrated generative AI capabilities of the *Collaborative Canvas*. We hosted hour-long sessions with groups of 2-3 participants in which we demonstrated how to interact with the canvas and invoke its generative AI capabilities. We then asked participants to use the tools to complete a short ideation task in which they each simultaneously wrote sticky notes manually on the canvas or used the AI tools to generate notes with as many ideas as possible to identify possible benefits of using LLM's in design thinking sessions. Afterwards they worked together to cluster ideas by theme.

Several important considerations arose from the feedback from our participants:

- Participants engaged with AI in varied ways. Some participants explored their own ideas first before incorporating AI-generated content; others leveraged AI support from the outset to either augment their understanding of the task domain or generate ideas.
- Most participants preferred to use the personal scratchpad to generate sticky notes rather
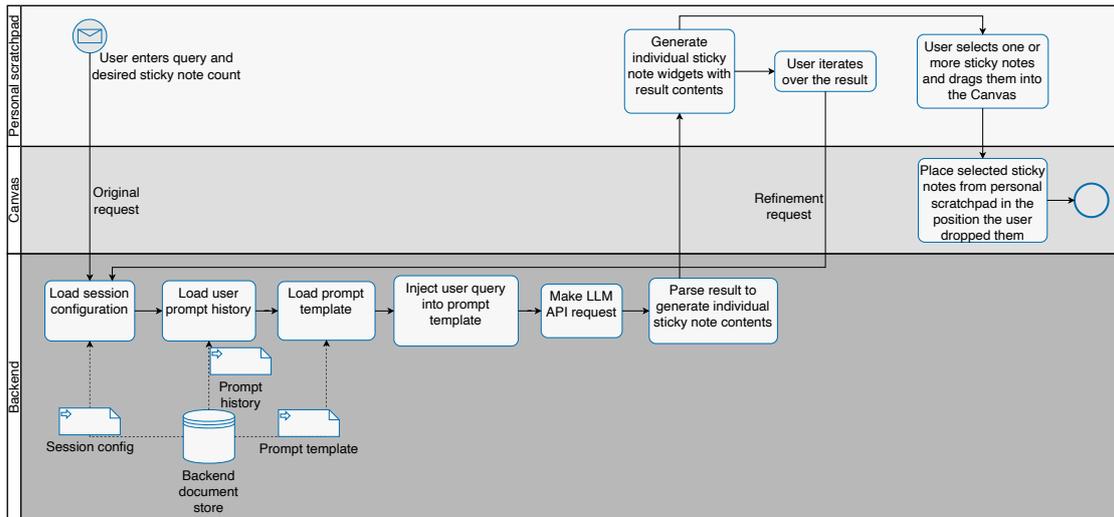
Figure 4: Sticky Notes Generation Flow. The diagram shows the activity flow for generating sticky notes. First, the user enters a query in their personal scratchpad. Next, the front-end UX calls an API endpoint in the back end, which looks up the user's canvas session information from the database. Then, if the user is making a refinement to existing sticky notes, the system retrieves the request/response history for those sticky notes. Next, the system loads the prompt template (Appendix A) and fills it in with the user's query and any retrieved refinement history. Finally, the system calls the LLM (via API) with the prompt to obtain a response, parses that response into a set of individual sticky notes, and adds them to the scratchpad UI.

than placing them directly on the canvas. Participants found that having this boundary between personal and shared work gave them control to review and filter the AI-generated outputs before making them visible to others.

- Some participants added AI-generated sticky notes to the canvas without modification, some participants modified them before adding them to the canvas, and others used them as inspiration for their own brainstorming.

- Participants felt that the generative capabilities could add value to their design thinking workshops by contributing novel ideas to get groups started in the brainstorming process, as well as to get groups "unstuck" when there was a lull in ideas.

- Participants identified that the role of the generative capabilities could extend beyond simply putting ideas on sticky notes. They envisioned that it could provide support for supporting their unique role as *facilitators* of a group brainstorming session, such as by clustering notes, keeping a workshop on schedule, and summarizing contributed ideas.

5. New directions

The development of the *Collaborative Canvas*, as well as our preliminary study, revealed several interesting new directions. We group these directions into specific improvements for the *Collaborative Canvas* tool and new research directions for human-AI co-creation.

5.1. Tool improvements

Differentiate AI-generated vs. human-generated outputs. Participants found it extremely important to distinguish between AI-generated and human-generated content in a group setting. Visual cues, such as icons and outlines, are one way to clearly represent which sticky notes were generated by the AI. Assigning credit for AI outputs may also involve capturing the prompt that resulted in the output, acknowledging the user who wrote the prompt and anyone who revised a sticky note, and displaying a record of both user and AI contributions for each sticky note. Rezwana and Maher [21] explored such issues of ownership of generated content and accountability for its use through a design fiction. They found that people's ethical stances on these issues depend on how the role of the AI is framed (e.g. as a tool vs. a collaborator). Given our participants' preference for reviewing AI outputs before sharing them, they may have viewed the *Collaborative Canvas* as a *tool* in which they maintain responsibility for AI outputs rather than a fully autonomous collaborator.

Provide control for the novelty and diversity of generated ideas. Participants recognized the need to sometimes produce unexpected (or provocative) ideas to push groups along their ideation process [22]. Although our tool did not expose any means for controlling low-level generative parameters (e.g. temperature), such controls could be provided. In addition, the model itself could be prompted to evaluate its own ideas for relevance to filter out ones that are less (or more) relevant to the current context.

Find better ways of hiding inference latency. In our preliminary study, we sometimes found that inference latencies were high, leading to situations in which a group of sticky notes were "pending" for long periods of time. These periods frustrated our users, especially when these sticky notes occluded other content. Future work is needed to design new mechanisms for hiding inference latency in a shared canvas environment.

5.2. New research directions for human-AI co-creation

Extend beyond content generation. Participants saw significant value in using AI to support their unique roles as facilitators of design thinking workshops. They envisioned helping groups get "unstuck" by generating new ideas. Facilitators also manage the brainstorming process by denoting different phases (e.g. idea generation vs. clustering vs. filtering & selection), and they envisioned that the AI could provide them with individualized support for organizing and executing these phases. Work by McComb et al. [23] identified four broad categories of AI roles within human-AI teams: as "tools," "partners," "analytics," or "coaches." The *Collaborative Canvas* acted as a tool by providing user-invoked content generation as a core function. Future work ought to examine how other roles might provide different forms of support to users beyond content generation, such as the facilitator role discussed by our participants.

Explore new interactive techniques. We explored invoking the generative capabilities of an LLM through direct manipulation interactions by generating sticky notes directly on the canvas or in the scratchpad. It is also possible to incorporate conversational interactions in the interface, such as by using a dedicated pane for Q&A (similar to the technique used by Ross et al. [10] in the Programmer's Assistant). Additional forms of interaction may also be possible. For example, could we generate a blending of two ideas by dropping one sticky on

top of another? Is there a way to "embed" conversation within the canvas? Given the rise of multi-modal models and the visual nature of the canvas, how can we incorporate images (or audio, or video) as outputs from (or inputs to) the generative model?

Evaluate human-AI team effectiveness. How does the use of generative AI in a group brainstorming activity influence the quality of the team's output? Does it make the process more efficient or require less mental effort? Further research is required to determine the precise quantitative impact of generative AI on group productivity.

AI proactivity. When should an AI assistant make a contribution to a group? Our tool must be explicitly *invoked* by a user in the group, suggesting a completely reactive mode of operation. However, mixed initiative interactions [24] may also be desirable, characterized by an AI agent that proactively generates content when it decides that such content may be beneficial to the group. How should the model determine when to proactively make a contribution? This decision is non-trivial; if the agent contributes the wrong content at the wrong time, it could be disruptive to the group's work process. New research is needed to assess when an AI agent should proactively contribute to a group's activity vs. when it should indicate that it has sometime to contribute vs. when it should remain silent and let the group focus.

6. Conclusion

We developed the *Collaborative Canvas* tool to explore group interactions with an LLM in the context of ideation tasks. The tool allows users to generate content by adding it to the canvas itself or by placing it in a private scratchpad that can be reviewed before it is added to the canvas. In addition, users can progressively refine content through context-sensitive re-generation requests. In a preliminary evaluation of the *Collaborative Canvas*, participants found utility in generated content by using it outright, modifying it, or using it to stimulate their own thinking. They also identified important considerations regarding the role of AI in an ideation session and made various recommendations for improving the tool. Our work motivates further study of AI assistance within group settings.

References

[1] A. B. Kocaballi, Conversational ai-powered design: Chatgpt as designer, user, and product, arXiv preprint arXiv:2302.07406 (2023).

[2] C. Yu-Han, C. Chun-Ching, Investigating the impact of generative artificial intelligence on brainstorming: A preliminary study, in: International Conference on Consumer Electronics - Taiwan, ICCE-Taiwan 2023, PingTung, Taiwan, July 17-19, 2023, IEEE, 2023, pp. 193–194. URL: https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226617. doi:10.1109/ICCE-TAIWAN58799.2023.10226617.

[3] J. Tholander, M. Jonsson, Design ideation with ai - sketching, thinking and talking with generative machine learning models, Proceedings of the 2023 ACM Designing Interactive Systems Conference (2023). URL: https://api.semanticscholar.org/CorpusID:259376359.

[4] B. Harwood, Chai-dt: A framework for prompting conversational generative ai agents to actively participate in co-creation, 2023. arXiv:2305.03852.

[5] D. Wang, E. Churchill, P. Maes, X. Fan, B. Shneiderman, Y. Shi, Q. Wang, From human-human collaboration to human-ai collaboration: Designing ai systems that can work together with people, in: Extended abstracts of the 2020 CHI conference on human factors in computing systems, 2020, pp. 1–6.

[6] C. Aragon, S. Guha, M. Kogan, M. Muller, G. Neff, Human-centered data science: an introduction, MIT Press, 2022.

[7] K. Inkpen, S. Chancellor, M. De Choudhury, M. Veale, E. P. Baumer, Where is the human? bridging the gap between ai and hci, in: Extended abstracts of the 2019 chi conference on human factors in computing systems, 2019, pp. 1–9.

[8] B. Shneiderman, Human-centered AI, Oxford University Press, 2022.

[9] I. Seeber, E. Bittner, R. O. Briggs, T. De Vreede, G.-J. De Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, et al., Machines as teammates: A research agenda on ai in team collaboration, Information & management 57 (2020) 103174.

[10] S. I. Ross, F. Martinez, S. Houde, M. Muller, J. D. Weisz, The programmer's assistant: Conversational interaction with a large language model for software development, in: Proceedings of the 28th International Conference on Intelligent User Interfaces, 2023, pp. 491–514.

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

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

[13] J. Koch, A. Lucero, L. Hegemann, A. Oulasvirta, May ai? design ideation with cooperative contextual bandits, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–12.

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

[15] C. Oh, J. Song, J. Choi, S. Kim, S. Lee, B. Suh, 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, 2018, pp. 1–13.

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

[17] A. W. J. Gan, Q. Dang, B. Western, J. L. G. del Castillo, Ai-mediated group ideation, eCAADe proceedings (2023). URL: https://api.semanticscholar.org/CorpusID:262061776.

[18] I. Cvetkovic, M. Gierlich-Joas, N. Tavanapour, N. Debowski-Weimann, E. A. Bittner, Augmented facilitation: Designing a multi-modal conversational agent for group ideation (2023).

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

[20] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini,

R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, T. Scialom, Llama 2: Open foundation and fine-tuned chat models, 2023. arXiv:2307.09288.

[21] J. Rezwana, M. L. Maher, User perspectives of the ethical dilemmas of ownership, accountability, leadership in human-ai co-creation (2023).

[22] S. M. Smith, N. W. Kohn, J. Shah, What you see is what you get: Effects of provocative stimuli in creative invention, in: Proceedings of the NSF International Workshop on Studying Design Creativity, 2008.

[23] C. McComb, P. Boatwright, J. Cagan, Focus and modality: Defining a roadmap to future ai-human teaming in design, Proceedings of the Design Society 3 (2023) 1905–1914.

[24] E. Horvitz, Principles of mixed-initiative user interfaces, in: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 1999, pp. 159–166.

A. Collaborative Canvas Prompt Template

```
1   This is a conversation with an automatic collaborative assistant that is expert,
2   eager, helpful, and humble. Here are some considerations the assistant must have
3   when generating the response to the user's query, which is located within the
4   <USERS_QUERY> tags:
5   - The assistant will respond in a way that the user can generate actions based on
6   assistant's response.
7   - The actions presented by the assistant should be in a numbered list format.
8   - The assistant will always indicate the end of the generated actions with the
9   string <END> at the conclusion of the list.
10  - The assistant should not use line breaks between each item in the numbered list.
11  - If the assistant receives context actions within the <CONTEXT_ACTIONS> tags, then
12  it must utilize those actions as context to generate the response actions.
13  - The assistant will respond with a maximum of 10 actions, but it can also respond
14  with fewer if the assistant deems it necessary or if the user asks it to generate a
15  few actions or a specific amount, such as a couple.
16  - The assistant must not generate actions with identical content.
17  - An example of the format that the assistant should use when responding is:
18  1. Response generated by the assistant for the first action.
19  ...
20  n. Response generated by the assistant for action number n.
21  <END>
22  <USERS_QUERY></USERS_QUERY>
23  <CONTEXT_ACTIONS></CONTEXT_ACTIONS>
```

Listing 1: Collaborative Canvas prompt template - The content within the <USERS_QUERY> and <CONTEXT_ACTIONS> tags is populated at runtime in the backend of the prototype once the user enters the query and the desired sticky note count, as shown in Figure 4.