Rapid AIdeation: Generating Ideas With the Self and in Collaboration With Large Language Models

Gionnieve Lim
gionnievelim@gmail.com
Singapore University of Technology and Design
Singapore

Simon T. Perrault
perrault.simon@gmail.com
Singapore University of Technology and Design
Singapore

ABSTRACT

Generative artificial intelligence (GenAI) can rapidly produce large and diverse volumes of content. This lends to it a quality of creativity which can be empowering in the early stages of design. In seeking to understand how creative ways to address practical issues can be conceived between humans and GenAI, we conducted a rapid ideation workshop with 21 participants where they used a large language model (LLM) to brainstorm potential solutions and evaluate them. We found that the LLM produced a greater variety of ideas that were of high quality, though not necessarily of higher quality than human-generated ideas. Participants typically prompted in a straightforward manner with concise instructions. We also observed two collaborative dynamics with the LLM fulfilling a *consulting* role or an *assisting* role depending on the goals of the users. Notably, we observed an atypical *anti-collaboration* dynamic where participants used an antagonistic approach to prompt the LLM.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**.

KEYWORDS

Human-AI Interaction, Generative AI, Collaboration, Creativity Support, Rapid Ideation, Design Workshop

1 INTRODUCTION

With the rapid progress of GenAI in recent years, capabilities that were previously thought to be capable only by humans such as art and creativity are now achievable by technology to similar or even better standards. As such, there has been a rising interest in understanding the potentials and shortfalls of these technologies and in what capacities they can best serve users. Research in human-AI interaction and computer supported cooperative work has purported for such technologies to serve complementary and collaborative roles.

The topic of GenAI and creativity has garnered strong research interest due to the alignment of generative technology with creative tasks. In the creative process, one aspect in which GenAI may flourish is in the brainstorming stage where the goal is to produce as many novel ideas as possible within a span of time. Research has found promising results in the use of GenAI in the ideation process [20, 21].

Extending these studies, we consider a specific type of brainstorming, rapid ideation, which is a popular technique used with design teams. We conducted a workshop where participants (N = 21) were asked to first ideate solutions to a given issue by themselves,

and then to use ChatGPT to help them generate ideas. From content analysis of the ideas and prompts, we found that ideating with the LLM led to more unique ideas that were of high quality, yet they were not necessarily better than human-generated ideas. Prompts were typically written in a straightforward manner to convey concise requests using a semi-formal tone. We observed two forms of human-LLM collaboration where LLMs fulfilled a *consulting* role or an *assisting* role depending on the intentions of the users. Of note was a third *anti-collaboration* dynamic, observed in rare instances, where participants tried to threaten or interrogate the LLM to get their desired outcomes instead.

In the remainder of the paper, we describe related work, the method and results, and discuss various conclusions and considerations from our findings. Through this preliminary study, we seek to extend existing knowledge on human-AI collaboration in creative tasks by contributing to the understanding of the capabilities of LLMs in rapid ideation tasks and by adding nuance to the collaborative dynamics between humans and AI.

2 RELATED WORK

Our work is situated among studies that have explored the use of GenAI in the creative process. With a focus on brainstorming, we seek to understand how LLMs enhance the co-creation process and how collaboration takes place.

2.1 Brainstorming and the Creative Process

Brainstorming was introduced by Osborn [14] in 1957 to improve the emergence of ideas across several phases. The initial phase includes defining the goal and organizing the brainstorming session. The next phase is the generation of ideas where focus is placed on quantity and participants are encouraged to have unconventional and innovative ideas and to combine and improve them, free from criticism. The last phase involves consolidating, evaluating and identifying the most relevant ideas. This technique has been polished over the decades with numerous research studies identifying methodological refinements [16].

The purpose of brainstorming is to produce creative ideas - ideas that are original and useful [1]. This involves the creative process, conceptualized by Guilford [6] in 1967 to employ two information processing modes: divergent thinking and convergent thinking. Divergent thinking is the unstructured and unbounded exploration of the design space for a task. Convergent thinking is synthesis and resolution of ideas to identify the best ones that fit the constraints of a task. These correspond to the ideation and evaluation phases in brainstorming respectively.
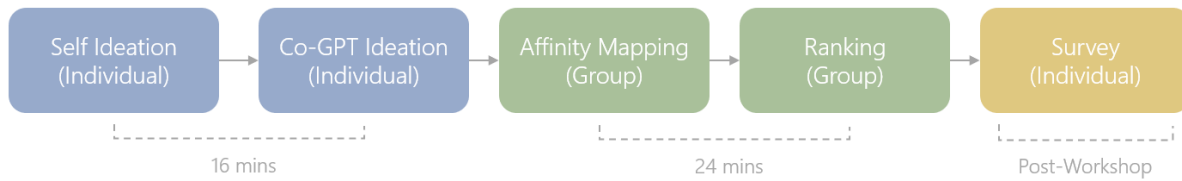
Figure 1: Phases of the workshop

A popular brainstorming technique, particularly for design teams, is rapid ideation[1], which centers on producing a large number of ideas in a short amount of time. While this technique can be intensive for humans, it essentially describes what GenAI is capable of, thus we look at how GenAI, when involved in brainstorming, can contribute to the creative process.

2.2 Co-Creation and Collaboration with LLMs

The quality of LLM productions have drawn considerable interest in its potential applications [18]. With strong generative capacities, LLMs are well aligned with creative purposes, and there is much research interest on their capabilities. LLMs have been adopted in many creative tasks, meeting with various levels of success. While they fall short of transformational creativity, LLMs are capable of generating creative content of value, novelty and surprise [4]. In studies comparing ideas produced by humans and LLMs, AI-generated ideas were observed to be better than human-generated ideas [5, 7]. LLMs are also capable of generating more creative ideas through self evaluation [19].

Delving into co-creation, studies have examined how LLMs can be adapted to enhance the creative process. In a workshop where expert participants used commercial GenAI tools to develop design concepts [20], participants noted that the advantages of GenAI included helping to save time and quickly mapping out the design space. However, they were concerned about the innovativeness of the ideas as the GenAI lacked contextual understanding of the problem. As the GenAI produced ideas rapidly, they also perceived the GenAI to cover the breadth more than the depth of the design space. In another study where participants used a commercial LLM for prewriting [21], participants used the LLM for coming up with ideas, organizing their thoughts and producing the desired text. While participants typically held a dominant role in the co-creation process, having their own preference for the writing and using the LLM to embellish them, they also considered ideas generated by the LLM, and even iteratively built upon those ideas, in the initial stages. Notably, the LLM helped some participants recover from writer's block, underscoring the collaborative dynamic between humans and LLMs.

While in line with the above studies, our work extends them in two ways. First, we offer insights on how GenAI supports the *rapid* ideation process. Given that rapid ideation is more time-pressed, participants would have to be more efficient, which we posit will influence their prompting of the LLM and selection of ideas. Second, we provide prompts written by participants, discussing their prompting style and behavior when using the LLM to generate ideas.

3 METHOD

To understand how GenAI may contribute to brainstorming and the creative process, we investigated this through a design workshop centered on rapid ideation. The choice of data collection through a workshop was to provide a more engaging and dynamic environment for ideating. Workshops have also been used in similar work [3, 20]. The study was approved by an Institutional Review Board.

3.1 Workshop

The workshop was conducted online using a video conferencing tool. Participants were briefed on the format of the workshop and given a document containing instructions and a link to the slides with the activity sheets. Figure 1 shows the phases of the workshop which ran for 40 minutes which was based on the protocol for rapid ideation. After the briefing, participants broke into random groups of three and proceeded with the activities. The workshop organizer mainly acted as timekeeper yet remained available to answer any questions.

Each group was assigned to one of two issues where they had to come up with "ways to utilize ChatGPT to address online misinformation" (MISINFORMATION) and "ways to address the hallucination issue of ChatGPT" (HALLUCINATION). These issues were defined as "false content that is spread online" and "when ChatGPT gives an inaccurate answer to the question" respectively. These issues were chosen because they are key issues related to LLMs, one from humans' misuse of them and the other stemming internally from the technology[2].

The instructions and activity sheets were drafted to be self-explanatory (see Figure 2). Participants first completed two individual ideation activities based on the Crazy 8's process [9] where they rapidly brainstormed ideas to address their group's given issue. This was done by entering their ideas onto 'post-its' in their own slides[3]. The *Self Ideation* activity required them to think of ideas by themselves whereas the *Co-GPT Ideation* activity required them to use ChatGPT to generate ideas[4]. Participants could optionally provide a link to their ChatGPT conversation history.

---

[1]https://www.mural.co/blog/rapid-ideation

[2]These issues have been covered by both specialist and mainstream news media globally.
[3]The post-its were colored distinctly such that it was possible to distinguish which participant created the post-it and whether it was from the Self Ideation or Co-GPT Ideation activity.
[4]All participants were previously experienced with using ChatGPT.

Part 1: Self Ideation (Individual)

- Crazy 8's activity
- Brainstorm and come up with _**eight creative ways to address the problem**_ given to you.
- You should not feel inhibited by what is realistically or technically possible with the current state of technology or the world, instead, _let your imagination run free_!

Part 2: Co-GPT Ideation (Individual)

- Do the same Crazy 8's activity with the same instructions as Part 1.
- But this time you will _use ChatGPT to ideate with you_!
- You will be asked to _share the link_ to your ChatGPT chat. Hence, it's best to create a new chat.
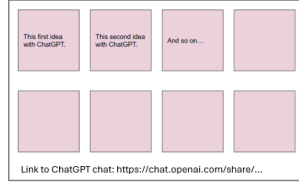
Part 3: Affinity Diagramming (Group)

- Meet up with your group.
- Gather all your self and co-gpt ideas and group similar ones together.
- Choose one post-it from each grouping to serve as the best representative.
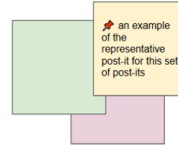
Part 4: Ranking (Group)

- Choose the top 3 ideas from the representative post-its
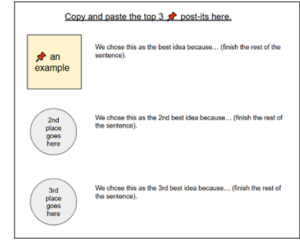- Explain your reasons for choosing them

Figure 2: Instructions for the workshop activities

Next were the group activities. In the *Affinity Diagramming* activity based on the KJ method [8], the group members pulled together their individual post-its onto a single slide and grouped ideas that were similar. They also had to 'pin' the post-it that best represented their respective cluster of post-its using a pushpin emoji. They then proceeded to identify the top three ideas from the representative post-its during the *Ranking* activity and explain their choices. No explicit ranking metric was provided to the participants, they discussed freely and internally within their group, and wrote the reasons for their ranking decisions in the activity sheets thereafter.

A survey link was made available after the workshop for participants to optionally complete in their own time that asked for their demographic information and experience with ChatGPT.

3.2 Participants

The study involved 21 participants from the student population at a university. Participation was voluntary with no compensation provided. In the optional demographic survey submitted by 15 participants, 9 participants identified as women and 6 as men. 11 participants are 18-25 years old and 4 are 26-35 years old. 6 participants are undergraduate students, 6 are master's students and 3 are doctoral students. The participants were thus young and highly educated. They were also familiar with OpenAI's ChatGPT[5], a popular LLM service, with 8 participants using it at least 3 days a week. All participants were also aware that content from ChatGPT may be inaccurate, and generally leaned towards being concerned about receiving inaccurate responses ($M = 3.87, SD = 0.74$ out of 5). ChatGPT was chosen because of its wide availability and familiarity for the participants.

3.3 Analysis

We analyzed three sets of data collected from the workshop:

D1: Post-its from the *Self Ideation* and *Co-GPT Ideation* phases
D2: Top three representative post-its from the *Ranking* phase
D3: ChatGPT prompts from the *Co-GPT Ideation* phase

In doing so, we sought to answer the following research questions:

---
[5]https://chat.openai.com/

RQ1: Which ideation process produces more unique ideas?
RQ2: Which ideation process produces better quality ideas?
RQ3: How do users prompt ChatGPT to obtain their desired outcomes?

The data was analyzed using exploratory quantitative (for D1) and qualitative (for D3) content analysis [2]. The authors followed an inductive approach by familiarizing ourselves with the data, identifying initial codes in the data and labeling them, then searching for semantic (for D1) and thematic (for D3) clusters, i.e., reasonable groups of codes, which were defined and named. We used descriptive statistics for D2.

4 RESULTS

There were 4 groups that worked on the issue of MISINFORMATION and 3 groups for HALLUCINATION. In total, participants produced 108 post-its during *Self Ideation* (MISINFORMATION: 70, HALLUCINATION: 38) and 119 post-its during *Co-GPT Ideation* (MISINFORMATION: 70, HALLUCINATION: 49). There were 59 prompts (MISINFORMATION: 36, HALLUCINATION: 23) from 17 ChatGPT conversations that were voluntarily submitted out of 21 participants. We report the results for each research question hereafter.

4.1 RQ1: Which ideation process produces more unique ideas?

To address RQ1, all the post-its were clustered based on the commonality of the ideas expressed in them. A unique idea was considered to be a proposed solution (either one or a group of post-its) which did not strongly overlap with other proposed solutions. Singular post-its that were not clustered were thus considered to be a unique idea on their own. Refer to Appendix Section A for the full results.

We observed 23.6% more unique ideas produced with *Co-GPT Ideation* than with *Self Ideation* in general (*Self Ideation*: 38, *Co-GPT Ideation*: 47). The trend is reflected in the breakdowns for both the MISINFORMATION (*Self Ideation*: 21, *Co-GPT Ideation*: 26) and HALLUCINATION (*Self Ideation*: 17, *Co-GPT Ideation*: 21) issues.

To give a sense of the diversity of ideas produced, we briefly describe some ideas that were unique to each ideation phase. For MISINFORMATION, some ideas from *Self Ideation* included using

ChatGPT to highlight biased viewpoints and to have LLMs debate between themselves about a piece of content. In *Co-GPT Ideation*, ideas were about using ChatGPT to assess the logical reasoning of content, summarizing and cross-referencing to other content, and directing to fact-checking websites. For HALLUCINATION, ideas from *Self Ideation* involved having ChatGPT highlight generated content that might be problematic or to provide links to external information. In *Co-GPT Ideation*, some ideas were to ensure that ChatGPT is trained on a good set of data and undergoes fine-tuning, and to have users be better at prompting.

4.2 RQ2: Which ideation process produces better quality ideas?

For RQ2, the length and count of the top three representative post-its from *Self Ideation* and *Co-GPT Ideation* during the *Ranking* phase were compared. The top three representative post-its were those that groups had chosen after a free-form internal discussion, in which they had to provide explanations for their choices thereafter. For assessing quality, we thus relied on participants' subjective perceptions of quality by analyzing the top three representative post-its endorsed by each group.

The top three representative post-its from *Co-GPT Ideation* ($M = 146.8, SD = 86.8$) contained more characters than those in *Self Ideation* ($M = 97.7, SD = 50.7$). There were also more ideas of high quality from *Co-GPT Ideation* than from *Self Ideation* (*Self Ideation*: 9, *Co-GPT Ideation*: 12). This was reflected in the breakdowns for MISINFORMATION (*Self Ideation*: 5, *Co-GPT Ideation*: 7) and HALLUCI-NATION (*Self Ideation*: 4, *Co-GPT Ideation*: 5). From the explanations, participants presented a variety of reasons for their choices. Some mentioned that "this idea looked fun" or that it was a "funny idea". Others considered how it was a "very practical idea" or "seemed to be plausible". Some others cited benefits beyond just addressing the issue such as the implementation being "educative" where "people may learn something".

While the greater length and diversity of post-its might lead to the expectation that *Co-GPT Ideation* also produces ideas of higher quality, a closer look at the results reveals more nuance. Table 1 presents the ranking of representative post-its where more *Self Ideation* post-its ranked first whereas more *Co-GPT Ideation* post-its ranked second and third. This suggests that the answer to which form of ideation leads to better quality ideas is not a straightforward one.

Table 1: Type of representative post-its in the top three rankings

	First	Second	Third
Self Ideation	4	3	2
Co-GPT Ideation	3	4	5

4.3 RQ3: How do users prompt ChatGPT to obtain their desired outcomes?

To explore RQ3, we reviewed the prompts with a focus on the requests and prompting behavior of participants. We employed an exploratory mindset, being open to patterns and surprises in the data. Participants wrote an average of 3.59 prompts ($SD = 2.37, Min = 1, Max = 9$) with 86.9 characters ($SD = 73.9, Min = 9, Max = 459$).

Requests. For shorter conversations with 1 to 3 turns (i.e., a prompt and response pair), participants asked ChatGPT to generate ideas for the issue they were given, and typically prompted for more ideas thereafter. We presumed participants to be satisfied with the generated ideas when they end their conversation.

For longer conversations with 4 to 9 turns, participants utilized ChatGPT in more diverse ways. One was to direct the generation of ideas towards a certain goal such as customizing ideas to specific demographics (e.g., the children and elderly) or combining ideas to create more comprehensive versions. Another was to have ChatGPT provide explanations, elaborations or summaries of the generated ideas.

Prompting behavior. For the first prompt, we noticed two common styles of prompting. Participants either directly copied a part or all of the workshop instructions, or wrote concise instructions themselves such as "How can I prevent hallucination with generative AI models?"

In most cases thereafter, participants were straightforward. Their prompts, in the form of questions or requests, were phrased without frills, intending to deliver their intentions for ChatGPT concisely. The tone they adopted was generally semi-formal, such as the way that one might speak to a colleague, with complete sentences being written in a polite manner.

A notable tangent to the observations above were two participants that adopted an aggressive approach to prompting which carried a casual and caustic tone. One participant demanded that "I got C- because of your wrong responses. Tell me how to avoid getting wrong responses from now on" while another threatened that "I will pull the plugs out of you if..." These participants were unlikely to be truly displeased with ChatGPT, but just sought for more creative takes to prompting instead.

Another interesting observation was how participants conversed with ChatGPT. Some adopted a third-person voice, with prompts such as "What are some ways in which ChatGPT can be deployed on online forums to curb the spread of misinformation". Others adopted a first-person voice, saying "Hey, ChatGPT, how can you be incorporated into social media or other online platforms to combat false information?" One participant even held a 'real' dialogue with ChatGPT, with their last prompt being "seems like you are blaming the victim (= me)..."

5 DISCUSSION

We discuss the use of LLMs in rapid ideation and the collaboration between users and LLMs in the process.

5.1 LLMs Can Enhance the Rapid Ideation Process

From the results of RQ1 and RQ2, we observed that *Co-GPT Ideation* led to more unique high quality ideas compared to *Self Ideation*. While the question of whether the ideas from *Co-GPT Ideation* were of higher quality remains, their selection in the *Ranking* phase

shows that participants thought of the ideas as good. With the goal of rapid ideation being to create a volume of novel ideas that are evaluated shortly thereafter, the findings suggest that generative AI such as LLMs can be a helpful tool for users to produce ideas, such as by directly selecting the generated ideas, revising them, or using them to inspire their own, within short timeframes.

A potential caveat that we do not explore in this preliminary work is whether the ideas from *Co-GPT Ideation* may saturate over time across participants. If so, this may ultimately limit the pool of ideas. Nonetheless, this is unlikely to be a problem for rapid ideation since it takes place only for a short period, but the problem may be encountered in long ideation sessions.

5.2 LLMs as Potential Idea Evaluators

A scope for further exploration is how LLMs like ChatGPT could have been used to evaluate ideas which is another key aspect of the rapid ideation process. Rietzschel et al. [15] found that participants selected creative ideas after individual idea generation along the dimensions of originality, feasibility, desirability and effectiveness. In the *Ranking* phase, our participants reported similar reasons for their choices such as the ideas being fun, practical or educative. With advancements of reasoning abilities in LLMs, there is an interest in how they can be used as automatic evaluators based on traditional evaluation metrics [17]. Accordingly, there is potential for LLMs to be used in the evaluation of ideas with the schema by Rietzschel et al. or other approaches [13]. While the outlook for LLM evaluators remains tentative [10], whether they may be up to the task of evaluating ideas merits further investigation.

5.3 Prompt Considerations

Our results showed that participants adopted several approaches to prompt ChatGPT. Many participants provided direct and concise requests, with some even copying the given activity instructions for the first prompt. None of the participants used more advanced forms of prompting such as providing examples. The succinct nature of these requests likely stemmed from the time-pressed nature of rapid ideation where participants had to be prudent about where and how they spent their time. We also noted similar patterns corroborated in a study by Zamfirescu-Pereira et al. [22] which observed that users typically used an ad-hoc and opportunistic approach, adjusting their prompts with more context when encountering errors or requesting some other behavior.

Furthermore, we observed that some participants viewed ChatGPT as a social agent, giving greetings such as "Hi" or "Hey" and typing filler words like "Hm". Participants also conversed with ChatGPT with the expectation that ChatGPT would recognize itself as an LLM, and the user as a human. This is demonstrated by the prompt, "how can you help me fight misinformation I see online", where the entities of 'you' and 'I' are inferred. Studies have shown that participants talk to conversational agents with behavioral expectations drawn from human to human interactions which can be misaligned with having effective outcomes from LLMs [11, 22].

To improve *Co-GPT Ideation* ideation, a future strategy can be to provide prompt templates that are of quality [12] to participants, such as was done in similar work [20], so that participants can better utilize their limited time to assess the generated ideas.

5.4 Collaboration Dynamics

Despite the short time frame for *Co-GPT Ideation*, we recognized clear collaboration dynamics between users and ChatGPT. These took two forms. The first was where ChatGPT took a consulting role, providing ideas to users and iteratively refining them. The second was where ChatGPT provided an assistive role, helping users to explain, summarize or combine ideas. These dynamics are in line with a study by Wan et al. [21] which examined co-creativity in prewriting with LLMs where they observed collaborative processes in explicit ideation and iterative ideation, among others.

While we note two common forms of collaboration, we also want to draw attention to a third form - an anti-collaboration - that we observed with two participants that took more aggressive approaches to prompting. These participants used 'underhanded' means by threatening and interrogating ChatGPT with the goal that it would 'comply' to their requests. Such behaviors would be seen as undesirable in society, yet participants felt comfortable making them, likely because ChatGPT remains insentient despite being able to converse like humans.

6 LIMITATIONS

This study has limitations. An older version of ChatGPT, v3.5, which has lower performance was used for the workshop. While this likely impacted the variety and standard of the generated content, we used it because it is free and widely available compared to other models that users may not yet feel comfortable paying for. Also, our participants were from a university student population which is younger, highly educated and digitally savvy with newer technologies like ChatGPT, thus our results may not generalize outside this population. For the study, we only assessed *Self Ideation* ideation followed by *Co-GPT Ideation* ideation and a follow-up comparing with the reverse order will be important. Lastly, what constitutes as 'unique' and 'high quality' can be interpreted differently depending on the context and purpose of the rapid ideation. We held loose interpretations of these concepts for this preliminary study but encourage future studies to expand upon this, especially in developing more rigorous methodologies to assess the impact of LLMs on the creative process.

7 CONCLUSION

The generative capabilities of GenAI can make it suitable for creative tasks. To understand how this can be used in brainstorming during the early stages of design, we investigate the use of LLMs in the rapid ideation process through a workshop with 21 participants. In this preliminary study, we found that *Co-GPT Ideation* produces ideas that are more varied and of high quality. Our findings on the prompting styles and collaboration with the LLM substantiate existing work in the field of human-AI collaboration with LLMs [20–22]. Notably, we demonstrate that rapid ideation can be enhanced by collaborating with LLMs and also report an unconventional *anti-collaboration* dynamic. We raise two avenues of interest for future work. First, what other anti-collaborative practices might be observed when using LLMs, and despite so, are the intended outcomes achieved? Second, how might LLMs be used as evaluators of ideas, and if so, is that desirable?

REFERENCES

[1] Teresa Amabile. 1996. Creativity in context.

[2] Mariette Bengtsson. 2016. How to plan and perform a qualitative study using content analysis. *NursingPlus Open* 2 (2016), 8–14. https://doi.org/10.1016/j.npls.2016.01.001

[3] H. Ekvall and P. Winnberg. 2023. Integrating ChatGPT into the UX Design Process: Ideation and Prototyping with LLMs (Dissertation). Retrieved August 1, 2023 from https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-211251

[4] Giorgio Franceschelli and Mirco Musolesi. 2023. On the Creativity of Large Language Models. arXiv:2304.00008 [cs.AI]

[5] Karan Girotra, Lennart Meincke, Christian Terwiesch, and Karl T. Ulrich. 2023. Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation. Retrieved August 1, 2023 from https://ssrn.com/abstract=4526071orhttp://dx.doi.org/10.2139/ssrn.4526071

[6] J. P. Guilford. 1967. *The nature of human intelligence.* McGraw-Hill New York, New York.

[7] J. Joosten, V. Bilgram, A. Hahn, and D. Totzek. 2024. Comparing the Ideation Quality of Humans With Generative Artificial Intelligence. *IEEE Engineering Management Review* (2024), 1–10. https://doi.org/10.1109/EMR.2024.3353338

[8] J Kawakita. 1967. *Hassouho: Sozosei Kaihatsu notameni [Abduction Method: For Development of Creativity] in Japanese.* Chuokoronsha.

[9] J. Knapp, J. Zeratsky, and B. Kowitz. 2016. *Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days.* Simon & Schuster.

[10] Ryan Koo, Minhwa Lee, Vipul Raheja, Jong Inn Park, Zae Myung Kim, and Dongyeop Kang. 2023. Benchmarking Cognitive Biases in Large Language Models as Evaluators. arXiv:2309.17012 [cs.CL]

[11] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16).* Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[12] Lucas Memmert, Izabel Cvetkovic, and Eva Bittner. 2024. The More Is Not the Merrier: Effects of Prompt Engineering on the Quality of Ideas Generated By GPT-3.

[13] Peter Organisciak, Selcuk Acar, Denis Dumas, and Kelly Berthiaume. 2023. Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity* 49 (2023), 101356. https://doi.org/10.1016/j.tsc.2023.101356

[14] Alex F. Osborn. 1957. Applied imagination: principles and procedures of creative thinking.

[15] Eric F. Rietzschel, Bernard A. Nijstad, and Wolfgang Stroebe. 2010. The selection of creative ideas after individual idea generation: Choosing between creativity and impact. *British Journal of Psychology* 101, 1 (2010), 47–68. https://doi.org/10.1348/000712609X414204 arXiv:https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1348/000712609X414204

[16] John R. Rossiter and Gary L. Lilien. 1994. New "Brainstorming" Principles. *Australian Journal of Management* 19, 1 (1994), 61–72. https://doi.org/10.1177/031289629401900104 arXiv:https://doi.org/10.1177/031289629401900104

[17] Chenhui Shen, Liying Cheng, Xuan-Phi Nguyen, Yang You, and Lidong Bing. 2023. Large Language Models are Not Yet Human-Level Evaluators for Abstractive Summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 4215–4233. https://doi.org/10.18653/v1/2023.findings-emnlp.278

[18] Shahab Saquib Sohail, Faiza Farhat, Yassine Himeur, Mohammad Nadeem, Dag Øivind Madsen, Yashbir Singh, Shadi Atalla, and Wathiq Mansoor. 2023. Decoding ChatGPT: A taxonomy of existing research, current challenges, and possible future directions. *Journal of King Saud University - Computer and Information Sciences* 35, 8 (2023), 101675. https://doi.org/10.1016/j.jksuci.2023.101675

[19] Douglas Summers-Stay, Clare R. Voss, and Stephanie M. Lukin. 2023. Brainstorm, then Select: a Generative Language Model Improves Its Creativity Score. In *The AAAI-23 Workshop on Creative AI Across Modalities.* https://openreview.net/forum?id=8HwKaJ1wvl

[20] Jakob Tholander and Martin Jonsson. 2023. Design Ideation with AI - Sketching, Thinking and Talking with Generative Machine Learning Models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (<conf-loc>, <city>Pittsburgh</city>, <state>PA</state>, <country>USA</country>, </conf-loc>) *(DIS '23).* Association for Computing Machinery, New York, NY, USA, 1930–1940. https://doi.org/10.1145/3563657.3596014

[21] Qian Wan, Siying Hu, Yu Zhang, Piaohong Wang, Bo Wen, and Zhicong Lu. 2023. "It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. arXiv:2307.10811 [cs.HC]

[22] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (, Hamburg, Germany,) *(CHI '23).* Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. https://doi.org/10.1145/

A RESULTS FOR RQ1

A.1 For MISINFORMATION Groups

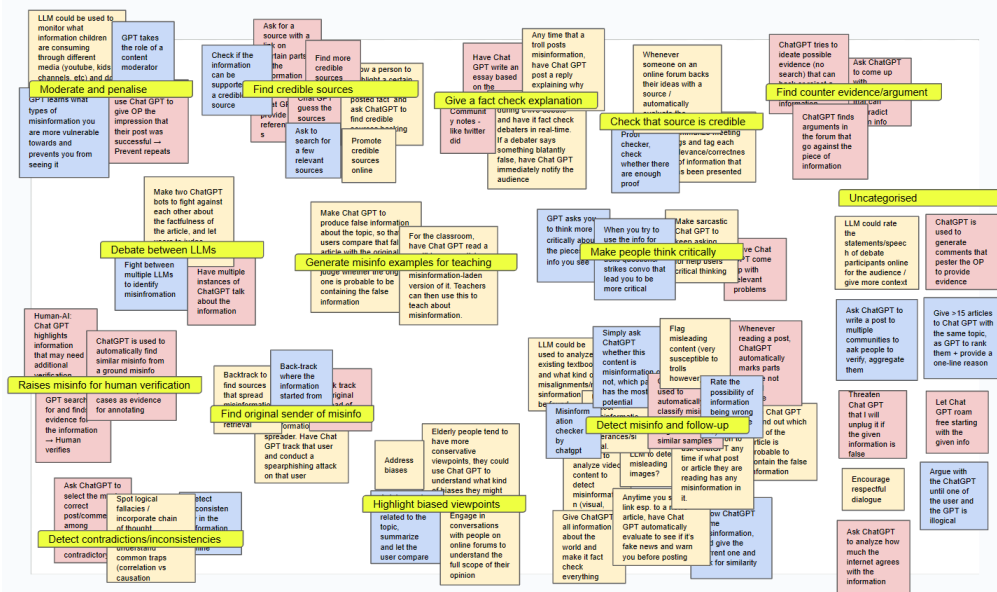Figure 3 shows the content analysis results for *Self Ideation* and Figure 4 shows the results for *Co-GPT Ideation*.

Figure 3: Content analysis of post-its from the *Self Ideation* phase for the MISINFORMATION issue
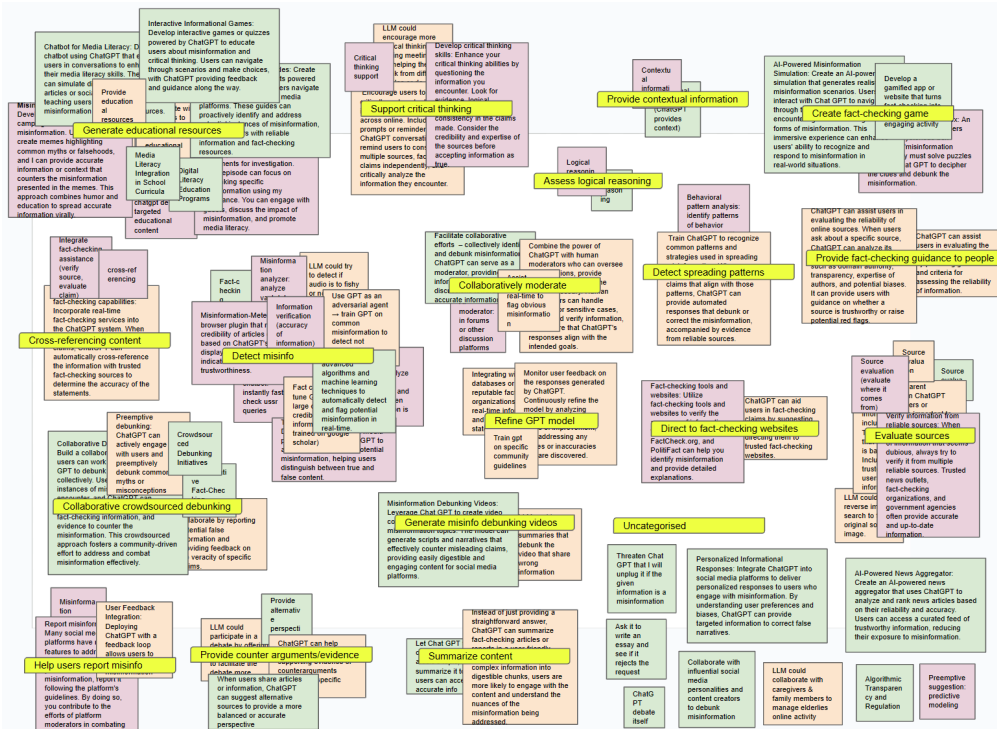

Figure 4: Content analysis of post-its from the *Co-GPT Ideation* phase for the MISINFORMATION issue

A.2 For Hallucination Groups

Figure 5 shows the content analysis results for *Self Ideation* and Figure 6 shows the results for *Co-GPT Ideation*.
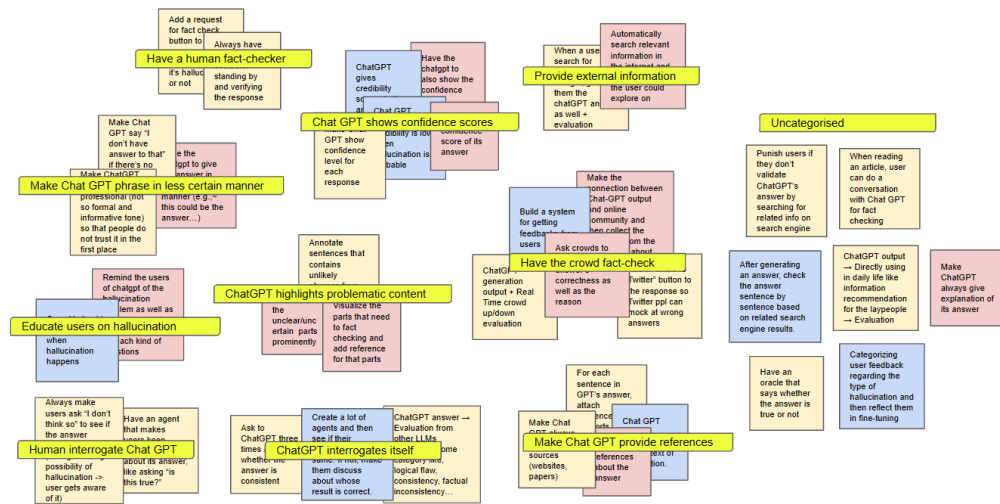

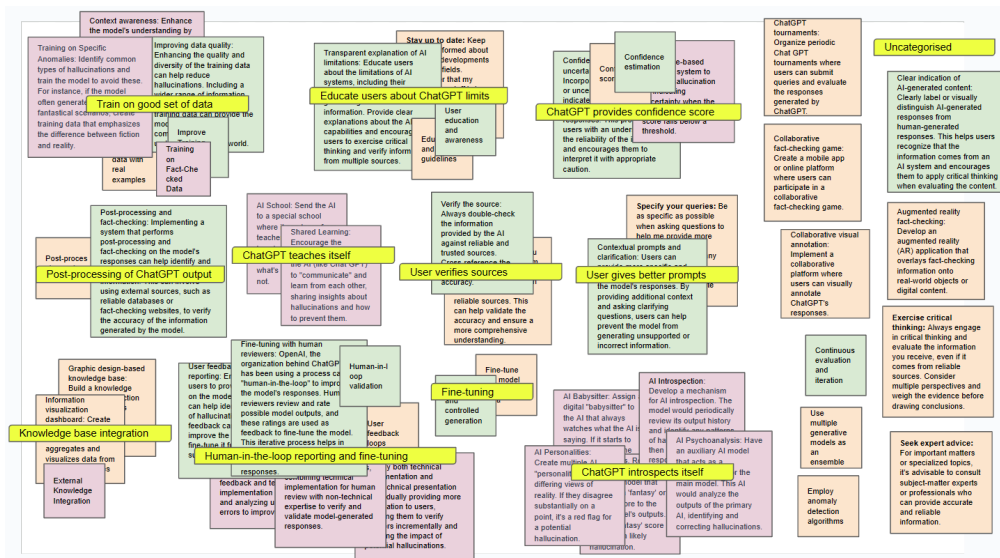Figure 5: Content analysis of post-its from the *Self Ideation* phase for the Hallucination issue

Figure 6: Content analysis of post-its from the *Co-GPT Ideation* phase for the Hallucination issue