When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task

Yuanning Han*
12010843@mail.sustech.edu.cn
Southern University of Science and Technology
Shenzhen, China

Ziyi Qiu*
213192675@seu.edu.cn
Southeast University
Chengdu, China

Jiale Cheng
19085542d@connect.polyu.hk
Hong Kong Polytechnic University
Hong Kong, China

RAY LC†
LC@raylc.org
City University of Hong Kong
Hong Kong, China

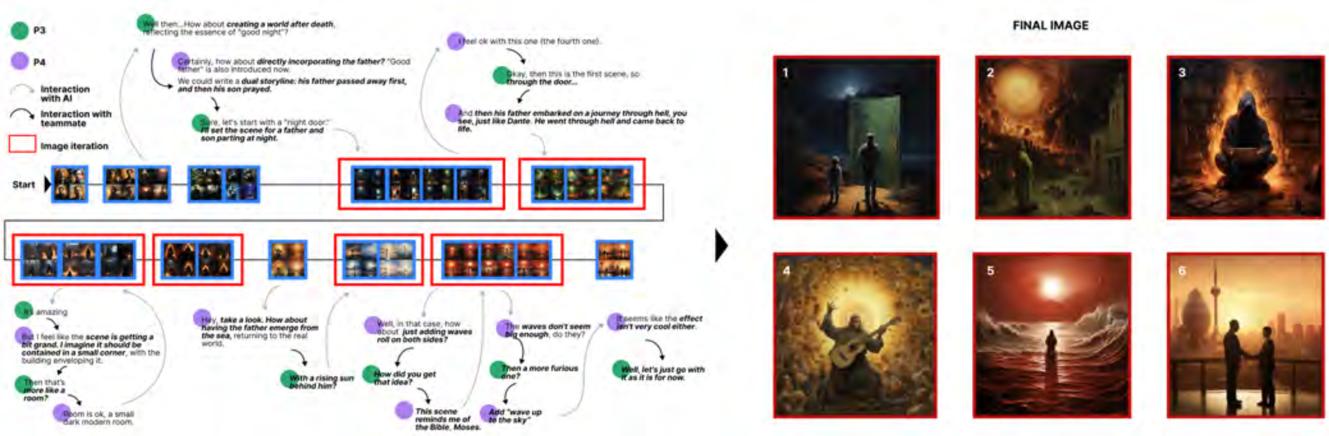

Figure 1: Two participants tried to use AI to draft the stage design based on a poem.

ABSTRACT

Studies of generative AI-assisted creative workflows have focused on individuals overcoming challenges of prompting for what they desire. How do collaboration and prompting influence each other, and how do users perceive the generative AI and their collaborator in the co-prompting process? We engaged students with design or performance backgrounds, with little exposure to generative AI in pairs, to design stages by prompting generative AI based on a poetic prompt. We found that participants followed patterns of either generating visual imagery for stories or vise-versa. The co-prompting encouraged participants to experiment and share their ideas, which stimulated discussion about creative content. Collaboration enabled participants to overcome the difficulties of collaborative prompting and reduced reliance on AI creativity, as participants valued the ingenuity of humans. This work highlights the importance of human-human collaboration when working with generative AI tools, suggests systems that take advantage of shared human expertise in the prompting process.

CCS CONCEPTS

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

KEYWORDS

human-AI collaboration, prompting in teams, GenAI, stage design

ACM Reference Format:
Yuanning Han, Ziyi Qiu, Jiale Cheng, and RAY LC. 2024. When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Hawai'i, USA.* ACM, New York, NY, USA, 15 pages.

*These authors contributed equally to this work.
†Correspondences should be addressed to LC@raylc.org.

1 INTRODUCTION

Generative AI (GenAI) has been employed in creative tasks wherein humans collaborate with AI in endeavors such as creative writing [15, 38], drawing [28, 30], performance [24], game design [18], and music arrangement [32]. AI-generated content (AIGC) can facilitate the human creative process by providing inspiration, offering novel ideas, facilitating human expression, and performing laborious tasks [15, 18, 43]. However, how does GenAI work with teams of human collaborators? Existing research primarily focuses on individual prompting with generative AI in creative tasks, which is a challenging endeavor, especially for non-expert single users [59]. Yet, the prompting process is vital to yield desirable outputs in text-to-image and text-to-text GenAI models. How do diverse human teams work together to overcome these challenges when interacting with the prompting process in creative collaborative tasks?

In addition to directly providing us with generative content for creative purposes, recent research shows that GenAI tools have the potential to facilitate decision-making [14, 45] and consensus-building [49] when they are involved in a collaboration scenario. A study found that compared with human teams that did not persist in enhancing their coordination over time, teams collaborating with AI were able to improve steadily [36]. However, further research into the detailed dynamics of team-AI collaboration is still needed, as it would inform us about how best to adapt to diverse voices in the collaboration process, developing strategies for optimizing decision-making, contribution, and efficient cooperation in teams that increasingly work together with GenAI systems.

Whether we can achieve the best performance when collaborating with AI depends on multiple factors, such as trust or acceptance towards AI [48]. For example, we can significantly develop users' trust and increase the acceptance of AI by increasing its interpretability or explainability [12, 13], thereby enhancing human-AI collaboration. Investigating a team's perception of generative AI performance can potentially lead to designs for more harmonious and efficient interactions, for example, conversational interventions or system mechanics for increasing the perception of reliability or transparency.

Team collaboration can be studied in multidisciplinary creative tasks like art design for stage performance [42]. Stage design involves artists, directors, designers, scriptwriters, and performers working in a cross-discipline environment where an artist talks with the performer, or performer with a designer, etc. Given that designs in this field are rooted in both visual and textual elements—with visuals often developed from texts like screenplay scripts, art design for stage performance provides a case study of collaborative practice where creative ideation is necessary. Inspired by both its collaborative characteristic and the possibility of utilizing generative AI during the creative process, we chose stage design as the task for our study.

We define co-prompting in this study (collaborative prompting) as a process that involves sharing and discussing prompts for generative AI systems among two or more individuals. Our study adopted a qualitative approach to investigating co-prompting in creative tasks and asked the following research questions:

RQ1: *What are the challenges for teams when co-prompting generative AI tools during creative design tasks, and what strategies do they employ to overcome them?*

RQ2: *How do individuals perceive the roles of GenAI and human collaborators in co-prompting generative AI for creative design?*

In this study, we conducted an online workshop where participants teamed up in pairs to design artwork for a stage performance. The design process involved co-prompting Midjourney based on inspiration from the poem "Do Not Go Gentle Into That Good Night," with the additional option of using ChatGPT (GPT-3.5). We used semi-structured interviews to probe each participant after the workshop and performed thematic analysis on the initial coding of this qualitative information to obtain insights into the co-prompting process.

The results from our study suggest that co-prompting enables participants to overcome challenges associated with building and adjusting prompt words. However, the prompting process remains challenging, and communication costs within the team can potentially increase. Nevertheless, co-prompting fosters a setting where participants feel encouraged to experiment and share their ideas, facilitating more in-depth discussions regarding the creative content and enhancing mutual understanding. This has proven beneficial for efficient and harmonious collaboration. However, discussions on prompting strategies could hinder meaningful conversations about creative ideas. Participants noted a shift in their roles towards guiding the AI and selecting its output, alleviating the need for laborious production. They also expected the generative AI to provide more unforeseen inspirations than desirable content that matches their imagination. During the workshop, participants actively sought their collaborators' opinions on both the prompting process and the generated output to bolster their trust in the GenAI system.

Armed with these results, we delve deeper into the challenges and strategies associated with using GenAI in collaborative scenarios and examine how users perceive it. We propose that the co-prompting process could be a double-edged sword for future collaboration, as it may facilitate creative ideation on the one hand to support team collaboration, but also add mental demands of strategic prompting, which could reduce team performance. Additionally, we propose that although users would sometimes fully trust and rely on GenAI to complete tasks, they still appreciate and even prioritize ideas from their own or their teammates, indicating the trust level within team-AI collaboration.

This study identifies challenges and strategies in humans working with other humans in the co-prompting creative process and explores how collaboration and prompting could mutually foster each other. We offer design considerations for developing collaborative creative design systems that cater to the needs of multiple users and suggest leveraging human-human collaboration in the GenAI creative process.

2 BACKGROUND

2.1 Human-AI Collaboration in Creative Process

Numerous projects in HCI research have integrated generative AI tools into the creative process. These studies suggest that AI has the potential to enhance the creative process, with many users reporting that AI often provides them with unexpected ideas [30]. This includes tasks such as deep comprehension and creative writing based on text [7, 21, 38, 58], mixed-initiative storytelling game [54], artistic visual creation [5, 8, 25, 43], and audio production [33, 44, 52], video generation [20], etc. Moreover, some research even explored the possibility of involving more than one generative AI agent in one creative task, such as using ChatGPT to generate a rich story and visualizing it with Stable Diffusion [22]. While the involvement of AI could give us more opportunities in the art domain, there are multiple challenges and strategies for prompt engineering during the creative process.

Within the context of using generative AI, most challenges were related to the prompting process, including how to build up and modify the prompts, especially for non-expert users who often struggle with how to get started or choose the right instructions [59]. Previous research has revealed that many artists desire additional assistance and guidance while constructing prompt words and refining the details [6]. While visual representations are especially needed in creative design, designers often face difficulties in the prompting process [60]. For example, it is still challenging to utilize Text-to-Image generative AI to translate one's original ideas into precise prompt words while ensuring richness, comprehensiveness, and accuracy, often resulting in images of random quality [57].

In terms of strategies, there were plenty of suggestions or practices for the improvement of prompt engineering and user interface. Prompt tools like "Promptify", which utilize a suggestion engine powered by large language models, can help users quickly explore and craft diverse prompts and improve the overall user experience with a clear interface [2]. Another tool designed by Singh's team can also assist with the prompting process through automatic prompt editing [50]. In addition to the improvement of AI tools, prompting strategies for common users such as "try multiple generations to get a representative idea of what prompts return" and "focus on keywords rather than the phrasings of the prompt" have also proven useful [31]. Seeking help on prompting from the web has also been observed as another common strategy for users to employ [59].

In sum, previous projects have indicated that users encounter challenges during the prompting phase due to inherent limitations of models and users' prompting abilities when using generative AI, and the main strategies focused on the improvement of AI tools and useful prompting skills for users. However, most of these findings and conclusions are based on a single person using generative AI. There is also a lack of understanding of the stage design-specific nuances of prompt creation. Building upon this, we intend to investigate what challenges and strategies would be like with the involvement and interaction of other users or what the dynamic of team-AI collaboration in creative process would be like.

2.2 Perceptions of AI in Creative Processes

Previous projects have highlighted the positive reception of generative AI during the creative process. Participants who often engaged in co-creation with AI consistently expressed positive experiences [43]. In some instances, they even described AI agents as their friends or co-workers instead of mere tools, emphasizing a noteworthy integration of AI into creative teams [10, 28]. Some individuals view AI as a creative partner, finding it to be an invaluable source of inspiration that enhances their overall working experience [18]. Besides, people cannot distinguish between AI and human-generated text, indicating that generative AI can effectively reproduce the nuance of the text written by humans [29].

Nevertheless, concerns about the risks and limitations related to generative AI exist. Beyond common risks such as bias, stereotypes, plagiarism, and diminished creative ideas, individuals have raised concerns about the role of generative AI, worrying about being replaced by generative AI tools in the future [38]. Meanwhile, researchers indicate that current tools still have their limitations and can only act as information providers instead of decision-makers [27]. Therefore, the predominant focus in process-oriented human-AI interaction research is on facilitating collaboration between humans and AI, rather than on exploring AI's potential as a replacement for human creativity [57].

It is noteworthy that the existing perceptions predominantly stem from individual interactions with generative AI. As we want to investigate the dynamics of team-AI collaboration, critical questions arise: How will people perceive AI when another human is in the team? Will there be a difference in collaborating with humans and AI during the creative process?

2.3 Human Collaboration in Stage Design

Since our research goal is to investigate the dynamics of team-AI collaboration, the interactions between human team members should also be considered. We have also investigated the dynamics in the creative collaboration process of human-only teams in stage design and identified several factors (cognitive diversity, shared mental model, and open communication) that could influence the experience and final performance.

For stage designers, effective teamwork and the designer's individual design skills are crucial throughout the design process. Stage designers are typically members of a design team, which may include the director, lighting designer, costume designer, sound designer, stage manager, music director, choreographer, and playwright or librettist [42]. The interdisciplinary nature of the design team stimulates profound thinking by enhancing cognitive diversity [39, 41]. Stage designers' work begins with a careful reading of scripts and identification of all the sensory elements needed for the story. This is followed by building rough models for discussion in meetings with the core design team [9]. During the meetings, designers with diverse experiences and knowledge need to reach a consensus on the general settings of the show, with sketches and drawings of detailed stage settings being needed for the show [9]. This co-construction process, involving collaborative creation of a shared understanding and a common output from individual ideas, has been identified as an effective method for enhancing team performance and satisfaction by developing a shared mental model

(SMM) within a team [34, 47, 56]. Additionally, open communication in this process can significantly enhance creative output [19, 55].

Yet, within this collaborative spectacle, challenges emerge. According to the interview with Emily McConnell, who has extensive experience working in a stage design team, issues such as how designers integrate their individual collaborative methods into a design team and how the design team as a whole works to translate the play from written words into visual art on stage are always the most obvious gaps and challenges during the collaborative process [35].

Reflecting on the potential of generative AI in the team-AI creative collaboration process and inspired by the creative teamwork during the stage design process, we decided to choose the stage design process (from script reading to conceptual sketching) as the creative task for our workshop to further explore the dynamics of team collaboration with generative AI in a creative process. We also want to investigate whether the discovered factors will be influenced by team-AI collaboration. It should be emphasized again that our focus of this research is still on team-AI collaboration, rather than the possibility of promoting further applications of generative AI in the stage design industry.

3 METHODS

3.1 Participants and Recruitment

We posted a call for participants on social media platforms and sent it via researchers' personal communication channels. Participants could join the workshop regardless of whether they had related knowledge of or experience in AI. We recruited 18 people (8 males, 10 females), all either holding or pursuing an undergraduate degree, with 12 out of 18 majoring in art and design (e.g., industrial design, product design, architecture design, painting), 3 of 18 having a background in performance (dancing), and one expressing an interest in pursuing stage design professionally, all participants had a design or performance background. Participants were randomly assigned into 9 pairs based on availability. All participants had an adequate command of English and were of Chinese ethnicity. They gave their consent to participate in the workshop and to have their data collected. All study procedures conformed to the institutional IRB guidelines on the human subject study, and the data collected was analyzed while being blind to the subjects' identities. Data were anonymized before analysis and deleted afterward. It is worth to noticing that we didn't require participants to necessarily have a background in stage design when recruiting for specific considerations. The primary objective of our study was not to examine the application of AI-generated content (AIGC) in stage design; instead, we aimed to investigate the effects of AIGC on team collaboration in creative tasks. So the reason why we selected stage design as our study's focus is that it could epitomize creative collaborative work. We also believe that our findings could have broader implications and are applicable to various forms of creative collaboration.

ID	*Group Number*	*Gender*	*Occupation*	*Art&Design Background*
P1	Group5	Male	Undergraduate	Architecture Design & Stage Design
P2	Group2	Male	Undergraduate	Digital Design
P3	Group7	Male	Undergraduate	Architecture Design
P4	Group7	Male	Undergraduate	Architecture Design
P5	Group3	Female	Undergraduate	Dancing
P6	Group1	Female	Undergraduate	Digital Art
P7	Group4	Male	Undergraduate	Industrial Design
P8	Group5	Male	Undergraduate	Industrial Design
P9	Group1	Female	Undergraduate	Industrial Design
P10	Group4	Female	Undergraduate	Painting
P11	Group2	Female	Undergraduate	Dancing
P12	Group3	Female	Undergraduate	Dancing
P13	Group6	Female	Undergraduate	Product Design
P14	Group6	Female	Undergraduate	Architecture Design
P15	Group8	Female	Postgraduate	Industrial Design
P16	Group8	Female	Graduate	Product Design
P17	Group9	Male	Graduate	Industrial Design
P18	Group9	Male	Undergraduate	Industrial Design

Figure 2: Demographic Information of Participants

3.2 Workshop Design

The online workshop (2 hours) asked participants to create five or more stage design sketches or references based on inspiration drawn from the Dylan Thomas poem "Do Not Go Gentle Into That Good Night." The workshop took place on the Tencent Meeting platform. To generate the sketches, participants had access to both ChatGPT (GPT-3.5) and Midjourney during the process (refer to Figure 3). Before each workshop, participants were given the text of the poem. Each workshop included a warm-up session to familiarize participants with text prompting, which is the only function that they could use when generating images on Midjourney during the workshop. They could generate as many images as they wanted but were asked to select five works as their design output. Screen sharing was needed when using generative AI, and the whole process was screen-recorded. Access to Midjourney and ChatGPT presents challenges in mainland China, as not all participants can access these platforms on their computers. If a participant was unable to access these platforms, they were offered remote control access through Tencent Meeting. On the other hand, if access was available, one participant in each group would use their computer to access both Midjourney and ChatGPT, sharing their screen during the Tencent Meeting. The participant tasked with operating ChatGPT and Midjourney was encouraged to actively listen to their collaborators' ideas during the prompting process. Due to potential network issues, participants may request to discontinue remote control access and instead ask the researcher to operate Midjourney and ChatGPT temporarily. In such cases, participants communicated their prompts to the researcher until the network problem was solved. Researchers observed the entire work process of the workshop, only intervening when necessary (e.g., technical assistance or answering questions). Participants were encouraged to think aloud and share ideas with teammates during the workshop. Considering the interdisciplinary nature of the discipline and the necessity of collaboration among individuals from different backgrounds in the design process, we chose stage design as our workshop task. We asked participants to digest and transform the inspirations from the poem into stage design sketches or references because stage design often takes inspiration from other art forms. We used Midjourney for text-to-image after pilot testing with other tools, due to its easy-to-use interface and ease of understanding for non-expert users. The way people work with AI and design in creative team collaboration is already demanding, so we aimed to minimize the additional cognitive load during other parts of the workshop.

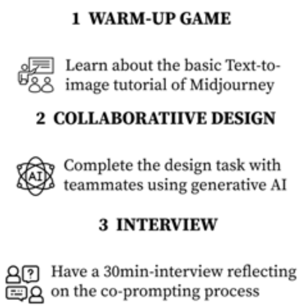

Figure 3: Three sessions in the workshop: (1)Warm-up game (2)Collaborative design (3)Interview

3.3 Interview Protocol

The semi-structured interviews were conducted remotely via Tencent Meeting with each workshop participant individually at the conclusion of each workshop. Each interview lasted between 20 and 30 minutes. Before the start of the interview, all participants were informed that their discussions would be recorded and transcribed. During the interview session, the researcher asked participants to recall and evaluate their experiences using GenAI in teamwork. They were prompted to identify challenges they faced in collaboration, describe their strategies to overcome these challenges, and express their feelings when encountering such issues. Additionally, participants were encouraged to assess the contributions made by GenAI and their human collaborators to their teamwork and to reflect on their own performance during the workshop. All interviews were conducted in Mandarin, subsequently recorded and transcribed using Tencent Meeting, and translated by the researchers into English.

3.4 Data Analysis

3.4.1 Semi-structured interview. The qualitative data obtained from semi-structured interviews were subjected to a systematic coding process using an inductive approach [3, 46]. Initially, two researchers independently familiarized themselves with the data by re-reading the interview transcripts [23]. This process allowed them to immerse themselves in the participants' responses and gain an understanding of the content. Following data familiarization, the researchers employed open coding to identify and label meaningful units of information within the transcripts [3]. Two authors who are native Chinese speakers independently conducted coding of the transcripts using an inductive approach. They individually analyzed the data and identified codes based on the content of the transcripts. Subsequently, they met to discuss any disagreements and worked towards reaching a consensus on the codes to ensure intercoder reliability. Codes were assigned to phrases, sentences, or paragraphs that captured key ideas, experiences, or concepts related to the research questions. The coding process was iterative, with the researchers continually refining and revising codes as new insights emerged [46].

To ensure consistency and enhance the reliability of the coding process, regular meetings were held between the researchers to discuss and compare their coding decisions [37]. Discrepancies were resolved through consensus, and any ambiguities or uncertainties were further clarified by referring back to the original data.

After the initial coding phase, the researchers engaged in a collaborative process to identify potential themes from the generated codes. They reviewed the codes and searched for patterns, similarities, and differences among them [17]. Codes that shared commonalities were grouped together to form preliminary themes. This thematic grouping was guided by the content of the data and the research objectives. The identified potential themes were then critically reviewed and refined. The researchers examined the relationships between the codes within each theme, ensuring that they were coherent and representative of the data. Themes were revised or combined when necessary to accurately capture the underlying patterns and concepts present in the qualitative data.

Throughout the coding process, researchers maintained detailed documentation of their coding decisions, including memos and reflective notes [11]. This documentation facilitated the traceability and auditability of the analysis, to maintain the rigor of the qualitative analysis.

3.4.2 Co-prompting process. A user journey map was employed as a data analysis method to gain insights into the experiences and interactions of a group of individuals while engaging with ChatGPT and MidJourney. One researcher reviewed all the recordings to collect all the prompting details during the experiment. Prompting processes were presented visually. The user journey analysis involved systematically examining the different touchpoints and stages that users went through during their interaction with GenAI systems. By mapping out the user journey, researchers were able to identify key moments of interaction, challenges faced, and valuable insights gained during the user's journey from text-based prompts to generated images [4, 51].

4 RESULTS

4.1 Collaboration Influences Co-prompting

4.2 Overall Strategies

When delineating the overall strategies adopted by participants, it is possible to thematically categorize them into two distinct groups based on their workflow (Figure 4). The first one, consisting of Group 2, Group 3, Group 8, and Group 9, adhered to a plot-based workflow. This approach necessitated the initial creation of a comprehensive narrative, with participants collaborating closely to construct a detailed storyline for the performance. The subsequent stage involved elaborating on the specifics of the stage scenes based on the established narrative. Conversely, the remaining groups - Group 1, Group 4, Group 5, Group 6, and Group 7 - employed a concept-based workflow. Here, participants initially concentrated on defining the thematic elements, style, or central

Figure 4: Two overall strategies observed during the workshops: (1)Plot-based workflow: the groups started from imagining the plot for stage, and then filled up the details and generated images in the following process (this workflow is employed by group 2, group 3, group 8 and group 9) (2)Concept-based workflow: the groups started from building up the concept for stage, and then generated images and completed the plot in the following process (this workflow is employed by group 1, group 4, group 5, group 6 and group 7)

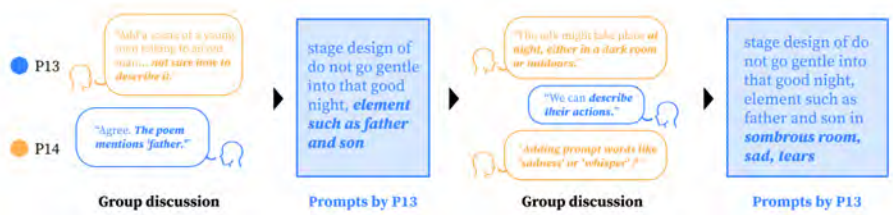

Figure 5: P13 and P14 tried to detail their conceptual ideas through co-prompting in Midjourney.

imagery for the stage, utilizing these foundational components as a springboard to craft the performance storyline and delineate the final scenes.

We noted that these two workflows seemed to be related to their professional backgrounds, and ChatGPT seemed to be more involved in Plot-Based groups during the collaboration process compared with Concept-Based groups. Moreover, as demonstrated in the following sections, we observed an interplay between co-prompting and collaboration throughout the collaboration process.

When co-prompting was introduced into collaboration, participants perceived that generative AI could swiftly produce concrete content representing their ideas, thus alleviating the psychological burden associated with experimenting with diverse concepts. Participants sought their collaborators' opinions in co-prompting, indicating that co-prompting may influence the mutual understanding within teams. This opinion exchange included discussions centered around prompt words and the sharing of ideas about generated images.

4.2.1 Co-prompting to develop vague ideas into prompt words. The co-prompting activities appeared to provide space for participants to collaboratively construct prompts, especially for some ideas that are still in the conceptual stage and challenging to articulate clearly. For example, we observed participants collaboratively fleshing out their thoughts through discussions regarding ideas from the poem that at first appeared to be simply inspiration ("sombrous", "sad") and developing into concrete prompts ("dark room," "whisper") (Figure 5).

4.2.2 Co-prompting to overcome the difficulties in adjusting the prompt. Participants often anticipate that GenAI will be able to comprehend their intentions in a manner similar to human understanding [59]. A significant number of participants highlighted the challenges associated with enabling generative AI to fully grasp their ideas, noting that it generally interprets only the literal meanings of the prompts, often failing to discern the underlying intentions or emotional nuances. P6 underscored this difficulty, stating, *"To utilize the AI properly, one needs to understand the correct methodology for instructing it to produce the desired results."* Participants found GenAI's limited capacity for association to be a hindrance, necessitating numerous iterations of clarifying and modifying their prompts to convey their ideas accurately.

To overcome the difficulties in articulating ideas to generative AI, participants collaboratively modified the prompt words, urging GenAI to deliver satisfactory output. They also speculated on how Midjourney interprets their prompts. When P13 and P14 prompted Midjourney to generate an image depicting a father and son in a sad atmosphere, they were dissatisfied with the generated results, so they tried refining their prompt words together (Figure 6).

Figure 6: P13 and P14 tried to modify the prompts together through multiple discussions and iterations

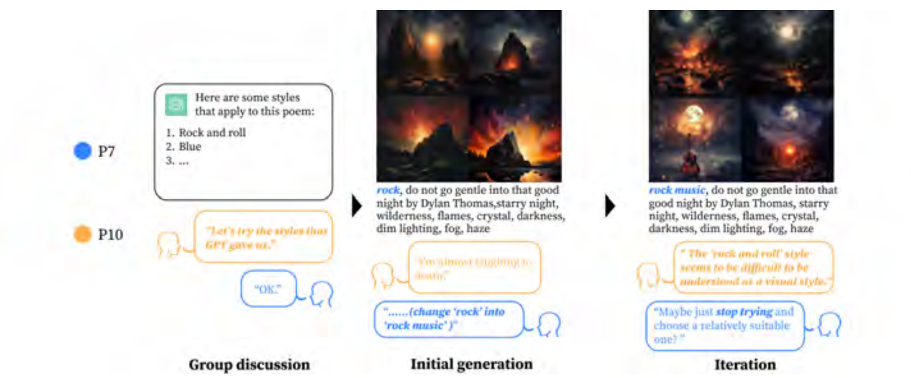

Figure 7: P7 tried to make the AI understand the meaning of 'Rock and roll' by changing the prompt word 'rock' into 'rock music', but eventually gave up.

4.2.3 Communication with AI can still be difficult even with the help of others. Participants reported increased difficulty in communicating in a way that both human collaborators and GenAI could understand one another, highlighting the challenges posed by the necessity to iterate prompts precisely within a group. Prompting the AI proved to be a particularly burdensome task that often discouraged further attempts at iteration. P7 reflected, *"The experience was discouraging, and I found myself tempted to abandon the task. The GenAI's lack of associative abilities made communication both draining and time-consuming, necessitating repeated clarification of our initial inputs."* (Figure 7). We also observed that participants would abandon a generation attempt after a few unsuccessful trials if they still could not achieve a satisfactory result. Moreover, participants noted a heightened sense of caution when prompting in a collaborative context to prevent errors that could potentially decrease efficiency and let the group down. P8 emphasized, *"In a co-prompting setting with others, I exercised greater caution to avoid mistakes and maintain efficiency."*

4.3 Co-prompting Also Influences Collaboration

During the workshop, participants collaboratively translated their concepts into prompt words and modified the prompt words when the previous generation was unsatisfactory. Although they were able to overcome the challenges together, the participants believed that co-prompting was still a challenging task because it increased the cost of communication in the team.

4.3.1 Co-prompting inspires discussion and leads to trying out of ideas. According to participants, GenAI appears to possess the capability to swiftly generate content replete with detailed representations derived from prompts, enabling participants to test their ideas in rapid succession and at low labor costs compared with manual production. This efficiency appears to encourage participants to explore and iterate their ideas with diminished psychological burdens. P3, a design student, noted, *"The manual drawing process is slow and laborious, particularly when revisions are necessary. GenAI facilitates easy adjustments, thereby fostering a richer ideation process."* This sentiment was echoed by P6, who emphasized the speed of Midjourney in image generation, allowing for *"low-cost trials and errors, and facilitating uninhibited idea validation."* Furthermore, P7 acknowledged the neutral position of GenAI in the collaborative setting, remarking, *"GenAI serves as a neutral third party on our team... we don't need to worry about its feelings."* Thus, the incorporation of generative AI into the team appears to encourage participants to experiment and ideate their ideas, aiding

Figure 8: (Above the dash line) P2 asked P11 for specific advice on prompt words before he actually typing them into Midjourney.(Below the dash line) P1 asked P8 for further advice before changing the prompts based on his own ideas.

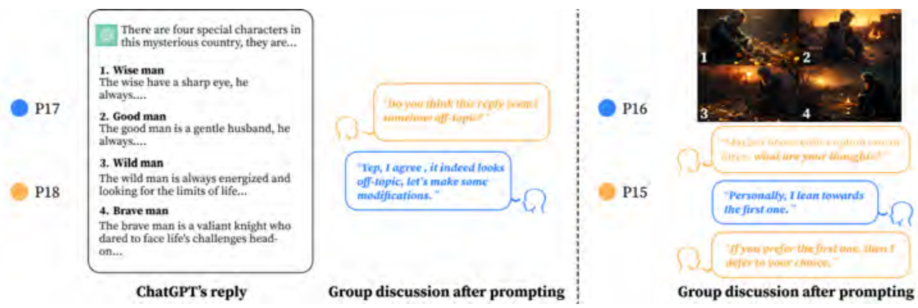

Figure 9: (Left column)P18 asked P17 about his ideas concerning ChatGPT's reply since P18 seemed not satisfied with it.(Right column)P15 asked P16 about P16's preference when selecting the images, and they eventually chose the one that P16 felt good about.

individuals in materializing and refining their concepts to achieve favorable results.

4.3.2 *Participants wanted to hear more from their teammates in co-prompting.*

During the prompting process, participants consistently sought to involve their collaborators, not only soliciting suggestions but also encouraging them to actively engage in evaluating and modifying prompts collaboratively. This was observed at various stages of the process, whether it was in the initial stages of crafting the prompt (Figure 8) or following the generated output phase (Figure 9). Throughout the workshop, participants undertook to further refine and optimize prompts based on the feedback received from their peers. This iterative collaborative process could cultivate outcomes that were met with satisfaction by all team members, as shown in Figure 10.

4.3.3 *Co-prompting can enhance mutual understanding in teams.*

We found that prompting GenAI collaboratively enhances the explicit information participants need to convey, facilitating the communication of concepts that might be challenging to express verbally. The end result is the materialization of abstract thoughts, contributing to understanding among human collaborators. Although GenAI sometimes gave rise to discrepancies in ideas from the original idea, participants believed that GenAI's contribution was beneficial in effectively conveying their core ideas. As P8 stated, *"While the information provided by AI wasn't fully aligned with my idea, it was sufficient to convey my thoughts."* In this context, generative AI appears to serve as a bridge, enabling a more tangible representation of participants' ideas and fostering mutual understanding in teams.

Participants found that co-prompting allows for discussions where ideas can be shared and assessed based on GenAI-created representations. P13 recalled an instance during her collaboration when her collaborator's interpretation of the AI-generated output changed her view: *"When we used Midjourney to generate this set of images, I initially thought the atmosphere did not align with the tone I interpreted from the poem. However, [P14]'s insights into the first image made me reconsider. Had I been working alone, I would probably have dismissed the set of images."* Moreover, in situations of divergent views, the GenAI-created outputs seemed to become a basis for negotiation, offering a more concrete understanding of each participant's perspective, and subsequently easing the path to reaching a consensus. P17 and P18, for instance, utilized ChatGPT to transform their individual ideas into stories. They then used these narratives as a means to deliberate on the feasibility of their ideas, a dynamic that not only increased mutual understanding but also played a key role in achieving a harmonious solution. (Figure 11).

4.4 Participants' Perception Towards GenAI and Humans in Collaboration

During the post-workshop interview, participants were asked to describe their perception of generative AI in the workshop process. Participants consistently viewed GenAI as a helpful agent in assisting them with tedious work. Participants selected GenAI-created

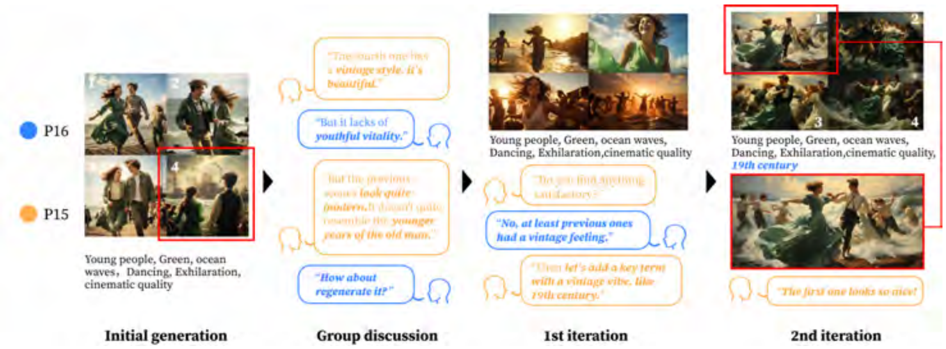

Figure 10: P15 and P16 tried to modify the prompts through discussions and iterations until both of them were satisfied with the results.

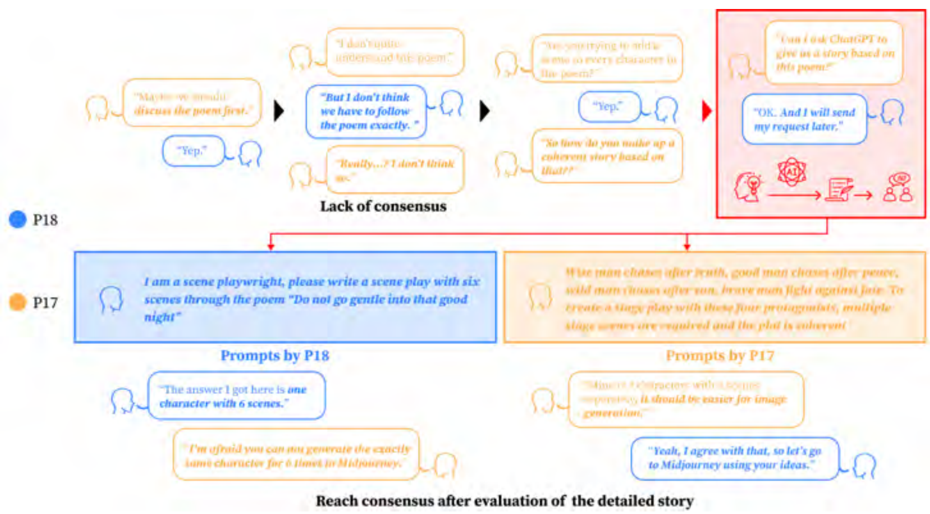

Figure 11: P17 and P18 initially disagreed with each other on several ideas, but they managed to reach the consensus after discussing ChatGPT's replies on detailed stories based on their own ideas.

outputs that aligned with their ideas while also expressing a desire for GenAI to provide them with unexpected inspiration. Although participants prompted GenAI to do creative work, they voiced their concern that human participants' creativity might be diminished by the GenAI workflow.

4.4.1 Human is responsible for guiding the GenAI to fulfill the laborious task. During the workshop, it was observed that several groups chose to assign creative responsibilities to GenAI, taking on roles similar to selectors who oversee, critique, and choose the GenAI-created outputs. The majority perceived generative AI as a tool rather than a collaborative partner, attributing this to the GenAI's incapability to actively engage in human discussions, comprehend emotional nuances, and partake in decision-making processes within the team. In contrast, P9 personified GenAI as a vital team member, noting its capacity to enhance productivity and foster creativity throughout the workshop. This sentiment was echoed by P17, who likened the GenAI to a consultant colleague offering valuable insights when humans got stuck coming up with ideas.

During interviews, participants described themselves in various roles, such as translators (P8), project managers (P8), film producers (P7), commissioners (P7), and design instructors (P7). They perceived GenAI through different lenses as well, categorizing them as designers (P8), listeners (P13), secretaries (P2), teammates (P9), service providers (P7), and design students (P7). Indeed, irrespective of the categorization, a consistent theme emerged: Participants viewed GenAI as primarily responsible for content generation and handling laborious tasks, while they steered the direction and provided guidance.

In particular, participants tended to view generative AI as a helpful agent that assists them, believing that it increases efficiency and saves labor in their collaborative work. They delegated tasks to GenAI over time, showing a shift in the distribution of labor. For example, after discussing their interpretation of the poem, P7 and P10 decided to use ChatGPT to extract pivotal words for further prompting. P7 explained, *"We used ChatGPT to save time and avoid having to read and interpret the poem over and over again. For me, GenAI is a time-saving and labor-saving tool."* Further instances of ChatGPT utilization encompass translation endeavors, materials

Figure 12: P13 and P14 were surprised by the 'broken moon' since it was not described in the prompts, but they seemed very satisfied with this unexpected result.

Figure 13: P3 and P4 were surprised by the generated images that Midjourney gave them, although it did not perfectly match with their original prompts. However, they still decided to choose it since the image surprisingly matched with the lines in the poem.

organization, and the extraction of design recommendations from pre-existing ideas or content.

4.4.2 Participants stated a preference for GenAI's output to match with their own, but also welcomed unexpected inspirations. After co-prompting GenAI, participants evaluated the outcome based on their preferences. They recognized the aesthetic quality of the generative AI output, interpreted the output, and selected the ones that better matched their expectations, iteratively revising the prompts. As P7 stated, *"We need to steer the GenAI in generating images and emphasize certain keywords."* However, participants also hoped the GenAI output could provide them with unforeseen inspiration. P8 suggested that "*You develop ideas based on your own experiences, which can lead you in a certain direction; AI could introduce you to a new way of thinking*." P17 added, *"I think the GenAI has its own preference due to the data they are trained, so the output generated by AI could offer us new inspirations."* Participants were often delighted to accept the ideas offered by GenAI. For instance, when Midjourney produced an image featuring a 'broken moon' (Figure 12), P14 expressed her surprise during the interview: "*When we think of the moon, we usually imagine a round, glowing orb. The GenAI's depiction of a broken moon was unexpected and intriguing ...it changed the atmosphere in a positive way. I appreciate these kinds of subtle touches*." Participants were generally willing to embrace unexpected content from the AI, even when it didn't precisely represent their original ideas (Figure 13).

However, when GenAI failed to deliver the unexpected insights participants hoped for, disappointment set in. P3 stated, *"I wish the GenAI could provide me with more 'aha moments'—something beyond what my prompts specify."* He added that if the GenAI couldn't provide additional inspiration, he would opt not to use it.

4.4.3 Participants perceived the original ideas of humans to be at a higher category than those of GenAI. Numerous participants with backgrounds in design underscored concerns about the potential diminishing of originality and creativity when incorporating GenAI into collaborative creative tasks. They at times expressed a preference for maintaining independence in their thinking devoid of GenAI. Echoing this sentiment, P1 expressed the desire to express individual emotion in creative endeavors, emphasizing, *"If I participate in the workshop again, I would want a human to interpret the poem, come up with subjective feelings... ChatGPT is powerful, but what it generated cannot represent my feelings."*

Furthermore, those well-versed in stage design voiced skepticism regarding the generative AIs' depth of understanding in the nuanced field of stage design. P1 found ChatGPT's insights into stage design somewhat rudimentary, while P14 criticized Midjourney's rigid approach to the discipline. Moreover, participants showed a tendency to value ideas from their human counterparts over suggestions given by the GenAI. Participants would invite their teammates to comment on and refine the generative outcomes instead of asking GenAI. This collaborative approach often led to more innovative results, showcasing a richer diversity of ideas. P15 observed, "Numerous suggestions from my teammates translated into outputs that were surprising and innovative."

Participants tended to trust their human collaborator more than the generative AI. They invited their collaborator to determine the validity of GenAI's output and trusted GenAI's output only if it aligned with human ideas, even if they were surprising ones. P17 mentioned, *"If I could find someone to check the output from GenAI with me, I would think the results given by GenAI are more reliable."* Simultaneously, some participants expressed appreciation for in-depth discussions on creative content, a facet that was somewhat overshadowed by the task emphasizing co-prompting strategies. Echoing this, P1 lamented, *"In this workshop, we paid excessive attention to how to use the tool and had less conversation on the creative content."*

5 DISCUSSION

5.1 RQ1: Challenges for Teams in Co-Prompting GenAI for Creative Design and Strategies for Overcoming Them

Co-prompting appears to be a double-edged sword for creative collaboration. Co-prompting enhances the mutual understanding between human collaborators and enables individuals to get help from their human collaborator during prompting. However, prompting generative AI remains challenging in the collaboration context, especially for non-AI expert users. Moreover, introducing GenAI in collaboration might induce participants to pay more attention to GenAI rather than to content creation.

5.1.1 Prompting remains challenging in a collaborative setting. During our workshop, participants encountered difficulties in verbalizing their thoughts, particularly in ensuring that both GenAI and their human collaborators understood their intentions. The incorporation of AI in the collaboration added complexity to the communication of thoughts. Furthermore, participants struggled to comprehend how GenAI would interpret prompts, particularly in text-to-image generation. At times, participants felt that GenAI grasped only the literal meaning of prompts, failing to capture the nuanced emotions and context they intended. This led to participants striving to adjust their prompts to enable the GenAI to understand their ideas.

Our findings indicate that participants are still applying a mental model of human-human interaction when interacting with AI, particularly with text-to-image AI. This observation aligns with Zamfrescu-Pereira et al.'s findings on why non-AI expert users face challenges when prompting Large Language Models (LLM) [59]. Participants expected the AI to understand the context of their prompts (e.g., interpreting "rock" as "rock and roll" rather than an actual rock). However, the text-to-image language model may interpret users' prompts more straightforwardly, without inferring the deeper, contextual, or metaphorical meanings that a human might naturally apply. This misperception towards the text-to-image model may lead participants to rephrase their prompts several times to achieve a preferable generation result, resulting in frustration after receiving undesirable outcomes (4.2.3). While iteration is considered a common strategy in prompt engineering [31], the repeated prompting by non-expert users in a collaboration context can be frustrating due to the lack of transparency in GenAI's mechanism.

Additionally, participants appeared to struggle with adjusting the prompt words, which may explain why some participants felt they were too focused on figuring out how to better use the generative AI rather than exchanging their creative ideas and having more in-depth discussions on them (4.4.3). In previous research, when the music composing process was intervened by generative AI, the participants also felt the GenAI might hinder the depth of their collaboration. They switched their focus to how to improve the content made by AI instead of having more creative engagement [53]. Besides, Zhang et al. found that a human-AI hybrid team has worse performance than a human-only team because the contribution of the AI lowers participants' mental demand and gives them an illusion of success, making the participants put less effort into the task [61]. The participation of AI in the co-creation process could distract participants from their creative interactions with team members, shifting their focus to how to prompt properly (4.4.3). Additionally, the output given by the GenAI may satisfy participants in terms of creating outcomes, making them less diligent in ideating for their own creations in the design process.

In summary, working with co-prompting in teams appears to be a double-edged sword, facilitating creative ideation on the one hand to support team collaboration, but adding mental demands of strategic prompting which could reduce team performance.

5.1.2 Strategies in collaboration could enhance prompting. Throughout the intervention, participants employed various strategies for prompting generative AI, including collaborative structuring of prompts and actively offering mutual suggestions. Additionally, the participants strategically assigned laborious tasks to GenAI. This approach allowed them to quickly test out their ideas and use the AI-generated outputs to materialize their abstract concepts, aiding in mutual understanding among human collaborators.

The collaboration enabled the participants to seek help in prompting from their human collaborators (4.2.1 and 4.2.2). Previous research has found that when non-expert users prompt GenAI tools individually, they seek help from online resources [59]. Besides, some artists actively looked for unique prompt words from various resources. This suggests constructing prompt words may require external assistance. The human participants in the workshop could serve as an "outer party" for their collaborators to request help in prompting.

Co-prompting is beneficial not only for enhancing the prompting process but also for enriching the collaborative experience. As one study notes, individuals tend to employ AI to handle large volumes of work when they emphasize productivity, thus reducing their own workload [1]. In the co-prompting setting, participants prompted GenAI to reduce their workload and save time in the team, assigning the more perceived burdensome tasks to the GenAI. Furthermore, participants utilized the outputs from GenAI as a visual medium to communicate their abstract ideas, facilitating mutual understanding between collaborators (4.3.3). Previous research has found that visual representation can reduce misunderstanding and conflict in collaboration [40]. In co-prompting, participants applied strategies that utilize generative AI to not only lighten their task load but also enhance their efficiency in communicating ideas.

5.2 RQ2: Individuals' perception towards the roles of GenAI and human collaborators in co-prompting

Our findings suggest that participants expect GenAI's output to fully represent their ideas, while they also welcome the unexpected inspiration delivered by GenAI. This may be due to participants' varying expectations of the roles of generative AI. Previous work has found that creative writers wanted to retain control over their writing strategies while co-creating with GenAI [1].We might expect designers to have similar expectations of GenAI when co-prompting with other designers, with GenAI respecting and adhering to the team's strategies in the design workflow. Indeed, our findings show that participants assigned GenAI to a subservient role in the design process (4.4.1) and categorized their strategies as more valuable than GenAI's (4.4.3), consistent with the findings on creative writers. In contrast, a study on designers suggested that they are more tolerant of GenAI's output accuracy when they view GenAI as a tool for inspiration [26], indicating a trade-off between expecting GenAI's support in a subservient role and its inspirational role in ideation. It appears that designers evaluate the ability of GenAI depending on the purpose for using it. Our study suggests that the role of the other human in the system falls into a hierarchy above the GenAI (4.4.3) even in the inspiration-generation phase.

Another concern we uncovered was participants' worry about the loss of creativity and lack of depth in thinking due to GenAI's intervention in teamwork, as expressed during the interview. Previous research suggested that GenAI can lower the mental demand of a high-performing human design team and make them explore less in the design task [61]. In our workshop, participants tended to follow the suggestions delivered by GenAI and missed chances to explore their design task further, which led participants to reflect during the post-workshop interview that they should not rely on the content provided by GenAI. Also, previous work showed that writers are more likely to be willing to collaborate with GenAI when they lack confidence [1]. Most of the participants did not have extensive experience in working with the discipline of stage design, which may also have caused them to lack confidence when working on design tasks, making them more receptive to suggestions from GenAI.

Our study observed participants utilizing plot-based and concept-based workflow depending on their background. We noted that participants with design backgrounds are more likely to apply the concept-based workflow (4.1). However, the plot-based workflow we observed is closer to the actual workflow of stage designers: they should read the script first and then identify all elements needed for the stage [9]. Additionally, previous research has shown that undergraduates in architecture and industrial design are more likely than design Ph.D. candidates to not plan the design process in advance [16]. So we speculate participants' behavior in the workshop could be influenced by their educational background.

One phenomenon we observed was participants' anthropomorphization of the GenAI (4.4.1). Some participants viewed GenAI in a more passive position (e.g., as a service provider), while others saw it in a more active role (e.g., as a teammate). Based on a previously published model, when individuals co-create with GenAI, they cycle between two states: highly engaging with GenAI and accepting its suggestions (Co-Creative Agentive Flow), or viewing GenAI as a tool to support their creation (Tool-Supported Creative Flow). [28]. Most participants perceived GenAI as guided by humans to create content, indicating a predominant experience of Tool-Supported Creative Flow in the co-prompting process. Nonetheless, they also accepted unexpected inspiration from GenAI throughout the workshop, showing that Co-Creative Agentive Flow exists in the co-prompting process. We speculate that the workshop design led participants to rely more on Tool-Supported Creative Flow: GenAI only responded when participants collectively decided to prompt it, positioning GenAI passively from the start. This argument is also supported by a participant (P18) who considers GenAI a tool that cannot actively join the discussion and decision-making process in their team. Furthermore, our observations indicated that participants tended to trust their human collaborators more than GenAI (4.4.3) and sought to have deeper discussions on creative ideas with humans collaborators only (4.3.2), suggesting a preference for more creative engagement with human collaborators over GenAI in co-prompting.

5.3 Design Implications

To optimize GenAI's effectiveness in creative collaborations, future systems should facilitate more natural, human-like interactions and ensure non-expert users have accurate perceptions of AI's capabilities. This way, users can focus more on actively discussing their creative ideas with human collaborators instead of being sidetracked by the complexities of using GenAI. Additionally, adapting GenAI to different stages of the collaborative design process, such as brainstorming and testing, is essential. For instance, during brainstorming, GenAI could produce more varied and unexpected outputs to spark inspiration, actively participating in discussions rather than merely responding when prompted. Conversely, in the concept validation stage, GenAI should more precisely mirror users' intentions if users want to visualize their ideas, aiding them in effectively communicating their thoughts to their human collaborators. An easy-to-use, scenario-specific GenAI system would enhance team creativity and efficiency, guiding them through challenging moments while honoring and embracing human ingenuity.

Furthermore, in designing future GenAI systems for collaborative creativity, it's crucial to consider the team's and individuals' performance to tailor interaction methods. Our analysis indicates that less confident users are more likely to embrace GenAI's suggestions, potentially leading to greater reliance on GenAI. However, users highly value their and their collaborators' originality. Therefore, the system should offer users more opportunities for reflection, aiming to decrease their dependency on GenAI.

5.4 Limitations

5.4.1 Online collaboration setting. One limitation is the exclusive use of an online collaboration setting, potentially influencing dynamics and outcomes compared to offline scenarios. When participants in our study worked with each other, usually one designer was the person entering the prompts while the other designer contributed. In real-world environments, this type of work is likely to have greater affordance for the other human when they are working offline together. This suggests that if the task is carried out offline, participants may value the human participants even further and further limit the role of the GenAI to be a support function. It is also possible that online and offline strategies are similar, given that one person would be in charge of the actual entering of prompt in both cases. However, we envision in offline cases that the prompt entering may be switched between the participants more frequently, giving the GenAI more of data-entry and output visualization roles.

5.4.2 Language and cultural nuances. The GenAI models we used for this study were mainly trained on English language data, while participants were native Chinese speakers, even though they performed the prompting in English. This mismatch may still introduce linguistic and cultural nuances impacting the interpretation and generation of AI-generated content. Exploring AI models specifically trained on Chinese language data could address this limitation and answer questions about whether the primary language of humans and GenAI must match for effective engagement to occur. This would also suggest that some of the differences in perception of the other humans and GenAI in our study could be due to subtler language differences.

5.4.3 Lack of control study of the interaction without GenAI. Our task put humans in pairs with GenAI to do co-prompting together. However, what would happen if no GenAI were present and the humans were simply designed as a team? Related works in studies of collaborative design have shown that in such interdisciplinary contexts, trust and communication are key themes of a varied style of collaboration across different contexts [40]. However, probing what happens when humans collaborate with humans would be reproducing existing literature to some extent, and does not tell us about the emerging case of AI-supported collaboration. Our study does not make a claim that co-prompting is *necessary* for any particular type of observed engagement, but rather that co-prompting can lead to particular types of engagement with, and perception of, GenAI. A study that does not use GenAI would not tell us how that engagement would go because it would not be studying a team's interaction with GenAI. Thus, we have focused in this study on the interesting patterns of interactions and perceptions that teamwork with GenAI can lead to.

5.4.4 Homogeneity of participant characteristics. Participants in the study shared similar demographic characteristics, such as age around 20 and educational backgrounds as university students, limiting the generalizability of findings to a broader population. It is possible that those of older demographics without as much experience with technology would perceive GenAI agents differently, such as treating them as experts, or rather mistrusting them. Future research should aim to include participants with diverse backgrounds and expertise to obtain a more complete understanding across different human participants.

5.4.5 Limited team size and generalizability. The small team size (two members) may not fully capture the complexities of larger teams, limiting the generalizability of findings to larger teams or real-world settings with more individuals. We may find that when dealing with larger teams, the human collaboration element becomes much more intertwined and difficult to disentangle on its own, and that collaborative prompting may be relegated to one individual. Thus differences in the team size may either enhance or reduce the reliance on prompting of GenAI. A proper study of the subject would involve giving team tasks that rely on multiple parties to accomplish and allowing them to work with GenAI either separately or with a single system.

5.4.6 Task and generalizability to different tasks. The use of a specific task (inspired by a poem to design for the stage) may not represent the range of tasks in human-AI teamwork. The findings and insights gained from our study should be interpreted within the context of a creative task that involves image generation based on textual inspiration. For example, a purely storytelling task may put the GenAI with a more equal footing with the human because they are both conversational agents. It may also create a greater divide between use of GenAI and teamwork because the task may be more engaging for the humans to discuss amongst themselves. Our study is thus specifically applicable to mostly text and image generation design tasks in pairs. However, we believe some of the results of GenAI perception and workflow processes may apply to design tasks in general, depending on the level of engagement.

5.4.7 Content analysis and analysis methods. In our studies, we mainly used qualitative approaches to analyze the data from the interviews and workshop recordings. Other dimensions of data that can provide further insight include the content of what is being produced, looking at specific properties like colors and geometric arrangements that may hint at how GenAI and humans effectively designed the outcome visually. We may also carefully look at the prompting text being used by individuals and compare the verbs and phrases that people used with each other during the engagement vs. those used to co-prompt with the machine. These different verbal behavior patterns may provide clues as to how people perceive collaboration vs. co-prompting in this task.

6 CONCLUSION

Our study explored the interplay between co-prompting and collaboration, and how teams perceive GenAI in the team-AI creative collaboration process. To directly observe the collaboration dynamics, we designed a workshop based on art design for stage performance and paired students and designers with performance backgrounds to work collaboratively with GenAI.

Our findings suggest that the involvement of GenAI in teamwork for creative tasks could be a double-edged sword for team performance. While GenAI can facilitate both the creative process and consensus-building, it can also introduce challenges in communication and strategic prompting. Additionally, we found that participants tended to prioritize ideas and suggestions from other human collaborators over those from GenAI, suggesting a hierarchy of priority within the team-AI collaboration.

This work underscores the significance of human-human collaboration when working with GenAI tools. While GenAI can enhance efficiency and provide inspiration, it appears to be most effective when integrated into a collaborative framework that harnesses the unique perspectives and creative abilities of human collaborators. These findings provide practical guidance to researchers and practitioners involved in the development of collaborative design tools in other creative domains. It highlights the value of understanding human-human collaboration in the context of generative AI.

REFERENCES

[1] Oloff C. Biermann, Ning F. Ma, and Dongwook Yoon. 2022. From Tool to Companion: Storywriters Want AI Writers to Respect Their Personal Values and Writing Strategies. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (, Virtual Event, Australia,) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1209–1227. https://doi.org/10.1145/3532106.3533506

[2] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 96, 14 pages. https://doi.org/10.1145/3586183.3606725

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[4] Snyder Carolyn. 2003. *Paper prototyping: The fast and easy way to design and refine user interfaces.* Morgan Kaufmann, Burlington.

[5] Eva Cetinic and James She. 2022. Understanding and Creating Art with AI: Review and Outlook. *ACM Trans. Multimedia Comput. Commun. Appl.* 18, 2, Article 66 (feb 2022), 22 pages. https://doi.org/10.1145/3475799

[6] Minsuk Chang, Stefania Druga, Alexander J. Fiannaca, Pedro Vergani, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. The Prompt Artists. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) *(C&C '23)*. Association for Computing Machinery, New York, NY, USA, 75–87. https://doi.org/10.1145/3591196.3593515

[7] Tsung-Shu Chang, Yitong Li, Hui-Wen Huang, and Beth Whitfield. 2021. Exploring EFL Students' Writing Performance and Their Acceptance of AI-Based Automated Writing Feedback. In *Proceedings of the 2021 2nd International Conference on Education Development and Studies* (Hilo, HI, USA) *(ICEDS '21)*. Association for Computing Machinery, New York, NY, USA, 31–35. https://doi.org/10.1145/3459043.3459065

[8] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. 2016. Empirically Studying Participatory Sense-Making in Abstract Drawing with a Co-Creative Cognitive Agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 196–207. https://doi.org/10.1145/2856767.2856795

[9] Christopher Dills. 2018. Read, render, realize. https://dramatics.org/read-render-realize/ Section: Backstage.

[10] Dina El-Zanfaly, Yiwei Huang, and Yanwen Dong. 2023. Sand-in-the-Loop: Investigating Embodied Co-Creation for Shared Understandings of Generative AI. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 256–260. https://doi.org/10.1145/3563703.3596652

[11] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of Advanced Nursing* 62, 1 (2008), 107–115.

[12] Andrea Ferrario and Michele Loi. 2022. How Explainability Contributes to Trust in AI. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1457–1466. https://doi.org/10.1145/3531146.3533202

[13] Luciano Floridi. 2019. Establishing the Rules for Building Trustworthy AI. *Nature Machine Intelligence* 1 (05 2019), 1–2. https://doi.org/10.1038/s42256-019-0055-y

[14] Andreas Fügener, Jörn Grahl, Alok Gupta, and Wolfgang Ketter. 2022. Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Info. Sys. Research* 33, 2 (jun 2022), 678–696. https://doi.org/10.1287/isre.2021.1079

[15] Katy Ilonka Gero and Lydia B. Chilton. 2019. Metaphoria: An Algorithmic Companion for Metaphor Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300526

[16] Gabriela Goldschmidt and Paul A. Rodgers. 2013. The design thinking approaches of three different groups of designers based on self-reports. *Design Studies* 34, 4 (2013), 454–471. https://doi.org/10.1016/j.destud.2013.01.004

[17] Greg Guest, Kathleen M MacQueen, and Emily E Namey. 2012. *Applied thematic analysis.* SAGE, California, CA, USA.

[18] Matthew Guzdial, Nicholas Liao, Jonathan Chen, Shao-Yu Chen, Shukan Shah, Vishwa Shah, Joshua Reno, Gillian Smith, and Mark O. Riedl. 2019. Friend, Collaborator, Student, Manager: How Design of an AI-Driven Game Level Editor Affects Creators. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300854

[19] Martin Hoegl and K Praveen Parboteeah. 2007. Creativity in innovative projects: How teamwork matters. *Journal of engineering and technology management* 24, 1-2 (2007), 148–166.

[20] Wenyi Hong, Ming Ding, Wendi Zheng, Xinghan Liu, and Jie Tang. 2022. CogVideo: Large-scale Pretraining for Text-to-Video Generation via Transformers. arXiv:2205.15868 [cs.CV]

[21] Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. 2022. Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. arXiv:2211.05030 [cs.HC]

[22] Ze Jin and Zorina Song. 2023. Generating coherent comic with rich story using ChatGPT and Stable Diffusion. arXiv:2305.11067 [cs.CV]

[23] Emma Jones, Sofia Torres, and Miguel Arévalo. 2020. Exploring the experiences of marginalized youth: A qualitative study. *Journal of Youth Studies* 23, 5 (2020), 678–695.

[24] Mirabelle Jones, Christina Neumayer, and Irina Shklovski. 2023. Embodying the Algorithm: Exploring Relationships with Large Language Models Through Artistic Performance. 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 654, 24 pages. https://doi.org/10.1145/3544548.3580885

[25] Pegah Karimi, Mary Lou Maher, Nicholas Davis, and Kazjon Grace. 2019. Deep Learning in a Computational Model for Conceptual Shifts in a Co-Creative Design System. arXiv:1906.10188 [cs.HC]

[26] Elisa Kwon, Vivek Rao, and Kosa Goucher-Lambert. 2023. Understanding inspiration: Insights into how designers discover inspirational stimuli using an AI-enabled platform. *Design Studies* 88 (2023), 101202. https://doi.org/10.1016/j.destud.2023.101202

[27] Yu-Ren Lai, Hsi-Jen Chen, and Chia-Han Yang. 2023. Exploring the Impact of Generative Artificial Intelligence on the Design Process: Opportunities, Challenges, and Insights. *Artificial Intelligence, Social Computing and Wearable Technologies* 113 (2023), 49–59.

[28] Tomas Lawton, Kazjon Grace, and Francisco J Ibarrola. 2023. When is a Tool a Tool? User Perceptions of System Agency in Human–AI Co-Creative Drawing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1978–1996. https://doi.org/10.1145/3563657.3595977

[29] RAY LC. 2022. Imitations of Immortality: Learning from Human Imitative Examples in Transformer Poetry Generation. In *10th International Conference on Digital and Interactive Arts* (Aveiro, Portugal, Portugal) *(ARTECH 2021)*. Association for Computing Machinery, New York, NY, USA, Article 8, 9 pages. https://doi.org/10.1145/3483529.3483537

[30] Yuyu Lin, Jiahao Guo, Yang Chen, Cheng Yao, and Fangtian Ying. 2020. It Is Your Turn: Collaborative Ideation With a Co-Creative Robot through Sketch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376258

[31] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 384, 23 pages. https://doi.org/10.1145/3491102.3501825

[32] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376739

[33] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376739

[34] John E Mathieu, Tammy L Rapp, M Travis Maynard, and Phillip M Mangos. 2009. Interactive effects of team and task shared mental models as related to air traffic controllers' collective efficacy and effectiveness. *Human Performance* 23, 1 (2009), 22–40.

[35] Emily McConnell. 2022. Designing Collaboration. https://plsn.com/archives/june-2022/designing-collaboration/

[36] Nathan J. McNeese, Mustafa Demir, Nancy J. Cooke, and Manrong She. 2021. Team Situation Awareness and Conflict: A Study of Human–Machine Teaming. *Journal of Cognitive Engineering and Decision Making* 15, 2-3 (2021), 83–96. https://doi.org/10.1177/15553434211017354 arXiv:https://doi.org/10.1177/15553434211017354

[37] Matthew B Miles, A Michael Huberman, and Johnny Saldana. 2014. *Qualitative data analysis: A methods sourcebook.* SAGE, California, CA, USA.

[38] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. 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 355, 34 pages. https://doi.org/10.1145/3544548.3581225

[39] Regina Moirano, Marisa Analía Sánchez, and Libor Štěpánek. 2020. Creative interdisciplinary collaboration: A systematic literature review. *Thinking Skills and Creativity* 35 (2020), 100626.

[40] Mimi Nguyen and Céline Mougenot. 2022. A systematic review of empirical studies on multidisciplinary design collaboration: Findings, methods, and challenges. *Design Studies* 81 (2022), 101120. https://doi.org/10.1016/j.destud.2022.101120

[41] Bernard A Nijstad and Wolfgang Stroebe. 2006. How the group affects the mind: A cognitive model of idea generation in groups. *Personality and social psychology review* 10, 3 (2006), 186–213.

[42] Berklee College of Music. 2023. Set Designer | Berklee. https://www.berklee.edu/careers/roles/scenic-designer

[43] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174223

[44] Zhaolin Qiu, Yufan Ren, Canchen Li, Hongfu Liu, Yifan Huang, Yiheng Yang, Songruoyao Wu, Hanjia Zheng, Juntao Ji, Jianjia Yu, and Kejun Zhang. 2019. Mind Band: A Crossmedia AI Music Composing Platform. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 2231–2233. https://doi.org/10.1145/3343031.3350610

[45] Charvi Rastogi, Yunfeng Zhang, Dennis Wei, Kush R Varshney, Amit Dhurandhar, and Richard Tomsett. 2022. Deciding fast and slow: The role of cognitive biases in ai-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–22.

[46] Johnny Saldana. 2015. *The coding manual for qualitative researchers.* SAGE, California, CA, USA.

[47] Catarina Marques Santos and Ana Margarida Passos. 2013. Team mental models, relationship conflict and effectiveness over time. *Team Performance Management* 19, 7/8 (2013), 363–385.

[48] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

[49] Joongi Shin, Michael A. Hedderich, AndréS Lucero, and Antti Oulasvirta. 2022. Chatbots Facilitating Consensus-Building in Asynchronous Co-Design. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 78, 13 pages. https://doi.org/10.1145/3526113.3545671

[50] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2023. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 68 (sep 2023), 57 pages. https://doi.org/10.1145/3511599

[51] Marc Stickdorn and Jakob Schneider. 2011. *This is service design thinking: Basics-tools-cases.* BIS Publishers, New York, NY, USA.

[52] Minhyang (Mia) Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 582, 11 pages. https://doi.org/10.1145/3411764.3445219

[53] Minhyang (Mia) Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 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* (, Yokohama, Japan,) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3411764.3445219

[54] Yuqian Sun, Xuran Ni, Haozhen Feng, Ray LC, Chang Hee Lee, and Ali Asadipour. 2022. Bringing Stories to Life in 1001 Nights: A Co-creative Text Adventure Game Using a Story Generation Model. In *Interactive Storytelling*, Mirjam Vosmeer and Lissa Holloway-Attaway (Eds.). Springer International Publishing, Cham, 651–672.

[55] Tang Tang, Valentina Vezzani, and Vikki Eriksson. 2020. Developing critical thinking, collective creativity skills and problem solving through playful design jams. *Thinking Skills and Creativity* 37 (2020), 100696.

[56] Piet Van den Bossche, Wim Gijselaers, Mien Segers, Geert Woltjer, and Paul Kirschner. 2011. Team learning: building shared mental models. *Instructional science* 39 (2011), 283–301.

[57] Mathias Peter Verheijden and Mathias Funk. 2023. Collaborative Diffusion: Boosting Designerly Co-Creation with Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 73, 8 pages. https://doi.org/10.1145/3544549.3585680

[58] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and Ray LC. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the IUI 2022 Workshops*, Vol. 10. CEUR-WS Team, CEUR, Helsinki, Finland, 56–65.

[59] 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/10.1145/3544548.3581388

[60] Chengzhi Zhang, Weijie Wang, Paul Pangaro, Nikolas Martelaro, and Daragh Byrne. 2023. Generative Image AI Using Design Sketches as Input: Opportunities and Challenges. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) *(C&C '23)*. Association for Computing Machinery, New York, NY, USA, 254–261. https://doi.org/10.1145/3591196.3596820

[61] Guanglu Zhang, Ayush Raina, Jonathan Cagan, and Christopher McComb. 2021. A cautionary tale about the impact of AI on human design teams. *Design Studies* 72 (Jan. 2021), 100990. https://doi.org/10.1016/j.destud.2021.100990