Decoupling of Usefulness and Novelty: Evaluating the Impact of Generative AI on Design Outputs and Novice Designers' Creative Thinking

Yue Fu
Information School
University of Washington
Seattle, Washington, USA
chrisfu@uw.edu

Tony Zhou
University of Washington
Seattle, Washington, USA
tyzhou05@uw.edu

Bin Han
Information School
University of Washington
Seattle, Washington, USA
bh193@uw.edu

Marx Wang
Information School
University of Washington
Seattle, Washington, USA
marxwang@uw.edu

Yixin Chen
Information School
University of Washington
Seattle, Washington, USA
yixin7@uw.edu

Zelia Gomes Da Costa Lai
Human Centered Design and Engineering
University of Washington
Seattle, Washington, USA
zelia@uw.edu

Rock Yuren Pang
Paul G. Allen School of Computer Science and Engineering
University of Washington
Seattle, Washington, USA
ypang2@cs.washington.edu

Katharina Reinecke
Paul G. Allen School of Computer Science and Engineering
University of Washington
Seattle, Washington, USA
reinecke@cs.washington.edu

Jacob O. Wobbrock
University of Washington
Seattle, Washington, USA
wobbrock@uw.edu

Alexis Hiniker
Information School
University of Washington
Seattle, Washington, USA
alexisr@uw.edu

Abstract

People are increasingly leveraging generative AI (GenAI) for design tasks, making it critical to understand GenAI's impact on design outcomes and users' creative capabilities. We conducted a within-subjects experiment where 36 participants designed advertisements both with and without GenAI. Evaluations from clients and online volunteers revealed that GenAI-supported designs were perceived as significantly more creative and unconventional. Additionally, online volunteers, but not clients, rated these designs as more visually appealing. However, neither group perceived differences in usefulness, and clients noted no improvement in brand alignment, highlighting a notable decoupling of novelty and usefulness (two established components of creativity) in GenAI-supported design outputs. Although short-term GenAI use did not broadly influence participants' creative thinking or experience, subgroup analyses indicated increases in divergent thinking among participants new to GenAI relative to participants with GenAI experience. We discuss the implications of the decoupling effect and GenAI's influence on humans' creativity.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Design, creativity, generative AI, human-AI collaboration

ACM Reference Format:
Yue Fu, Tony Zhou, Bin Han, Marx Wang, Yixin Chen, Zelia Gomes Da Costa Lai, Rock Yuren Pang, Katharina Reinecke, Jacob O. Wobbrock, and Alexis Hiniker. 2026. Decoupling of Usefulness and Novelty: Evaluating the Impact of Generative AI on Design Outputs and Novice Designers' Creative Thinking. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3772318.3791329

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791329

1 Introduction

Creativity, commonly defined as the ability to generate novel and useful ideas [102], is increasingly recognized as a pivotal skill in modern society. Research has highlighted creativity's significant role in improving problem-solving, fostering innovation, boosting productivity, and contributing to personal socio-emotional well-being, satisfaction, and life success [6, 76, 93, 96, 97], noting its significance for both individual development and societal advancement.

The design industry is a key environment for applying creativity, where the constant generation of fresh and innovative ideas is essential [24]. Creativity drives the production of novel and useful outputs [10] that meet client needs and enable businesses to stand out in competitive markets. In 2021, the global design industry was estimated at $162 billion and consistently growing [38].

To address the increasing demand for innovative design solutions, creativity support tools (CSTs) have become widely used in the design. For example, over 90% creative professionals worldwide use Adobe Photoshop, which boasts an estimated 23 million monthly users [59]. Canva, an online graphic editing tool launched a decade ago, has attracted 170 million monthly users worldwide in early 2024 [16]. These CSTs are designed to enhance both the creative person and the creative activity [41], aiming to "make people more creative more often" [90] and facilitating creative activities by improving the processes, usability, and workflows involved in creative work [83].

Recently, GenAI-powered CSTs have gained significant traction for their potential to transform how designers work. These GenAI tools claim to accelerate the creative process by automating repetitive tasks [7], suggesting innovative design concepts [105], and providing real-time feedback [109]. Text-to-image models like OpenAI's DALL-E [69] and Midjourney [67] can generate high-quality visual content from text descriptions, enabling designers to quickly prototype ideas [28]. For example, Adobe's Creative Cloud suite has incorporated GenAI features like "*Generative Fill*" [1] and "*Neural Filters,*" marketed for their ease of use[2].

However, as the design industry increasingly relies on these tools, the implications of these tools on users' creative experiences and the resulting design outputs remain to be studied. Our research seeks to answer two questions:

- **RQ1:** How do generative AI tools influence users' experiences and creativity during a design task?
- **RQ2:** How do generative AI tools influence the perception and evaluation of the resulting design outputs?

To investigate these questions, we conducted a two-phase study (See Figure 1 for the study overview). In a within-subjects lab experiment with 36 participants from diverse design backgrounds, participants were asked to complete a common design task with or without GenAI support: creating advertising material for a social media campaign to promote a client organization. We measured participants' mood, creative ability, and self-reported self-efficacy. Each participant submitted one or two designs per session and a total of 105 eligible designs were collected.

In the second phase, the collected designs were evaluated by five people from the organization the design task targets ("*clients*") and 155 independent online volunteers, all blinded to experimental conditions. Evaluations were rated on visual appeal, creativity, unconventionality, and perceived usefulness. The clients also rated their perception of the designs' alignment with the client's organization. Clients also provided optional qualitative feedback on each design.

Our study reveals that a single 25-minute session with GenAI showed no significant impact on participants' inherent creative abilities, mood, or self-efficacy. However, subgroup analyses indicate that the effects of GenAI varied based on prior GenAI exposure: participants new to GenAI exhibited increased divergent thinking compared to those already familiar with these tools.

Regarding design output, we discover that GenAI-supported designs are consistently perceived by both clients and online volunteers as significantly more creative and unconventional compared to human-only designs, which aligns with recent studies on GenAI's influence on design outcome [18, 19, 110]. **However, we also find, surprisingly, these GenAI-supported designs are not perceived as significantly more useful, showing a critical decoupling of novelty and usefulness (two traditional dimensions of creativity) of design output. The finding complicates prevailing optimism about GenAI's role in creative design, suggesting that increased novelty does not necessarily translate into more effective design outputs.** In addition, qualitative analysis of client feedback further illuminates this decoupling effect, with GenAI-supported designs often praised for their innovative and colorful qualities but criticized for"*too busy*" or deviation from social norms.

Our study provides empirical evidence of GenAI's influence on users and design outputs. We discuss the observed decoupling effect GenAI has on traditional two components of creativity: novelty and usefulness. We also discuss this effect on creativity and society as a whole. In addition, we argue for the intrinsic value of creativity, highlighting the importance of designing GenAI tools that support human creativity rather than diminish it through automation.

2 Related Work

2.1 Creativity: Aspects, Measurement, and Enhancement

The exploration of human creativity has intrigued scholars since ancient times [47]. One famous framework, the 4P theory of creativity, delineates creativity into four aspects: *Person* (individual traits like personality, cognitive abilities, and motivation), *Process* (stages involved in creative thinking such as ideation and evaluation), *Product* (the outcome of creativity marked by originality and impact), and *Press* (environmental factors influencing creativity) [85]. Our research primarily focuses on evaluating GenAI's impact on the *Person* and *Product* aspects of this framework (see Fig. 2).

Creativity from the *Person* perspective involves key components such as divergent and convergent thinking [108]. Divergent thinking is defined as the measure of the varied potential of creative thought and focuses on the processes of creative ideation [86, 89]. Divergent thinking relates to the ability to generate multiple creative ideas, typically assessed via tests like the Alternative Uses Test (AUT) [35] and Torrance Tests of Creative Thinking [17, 95], which measures how creatively participants can think of non-common uses for common objects like a "paper clip" [35, 53]. The AUT evaluates a person's creativity across three facets: *fluency*, the quantity of ideas generated; *flexibility*, the variety of idea categories; *originality*, uniqueness or novelty of the ideas are compared to the ideas generated by other participants. On the other hand, convergent thinking involves narrowing down these possibilities to find a single correct

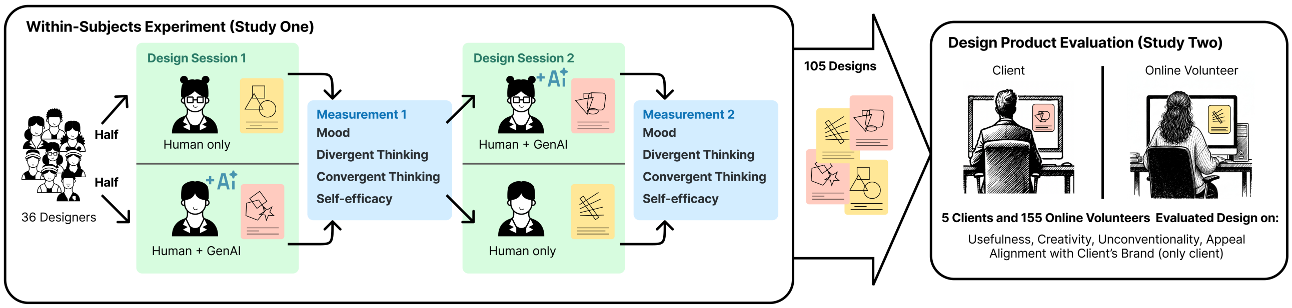

Figure 1: Our within-subjects experiment (Study One) and design output evaluation (Study Two) procedure. We recruited 36 participants to create social media advertisements promoting a client's organization. Participants were randomly assigned to two groups and completed two design sessions in a counterbalanced order. In the first session, one group designed independently (human-only), while the other collaborated with a GenAI tool (GenAI-supported). After the first session, researchers measured participants' mood, self-efficacy, divergent thinking with the Alternative Uses Test [4], and convergent thinking with the Anagram Test [12]. In the second session, the groups switched conditions: the human-only group used the GenAI tool, and the GenAI-supported group designed independently. The same measurements were conducted after the second session. Across both sessions, a total of 105 eligible designs were collected. In Study Two, five clients and 155 online volunteers, blind to the experimental conditions, evaluated the designs based on their usefulness, creativity, unconventionality ("weirdness"), alignment with the client's brand (only evaluated by the clients), and visual appeal.

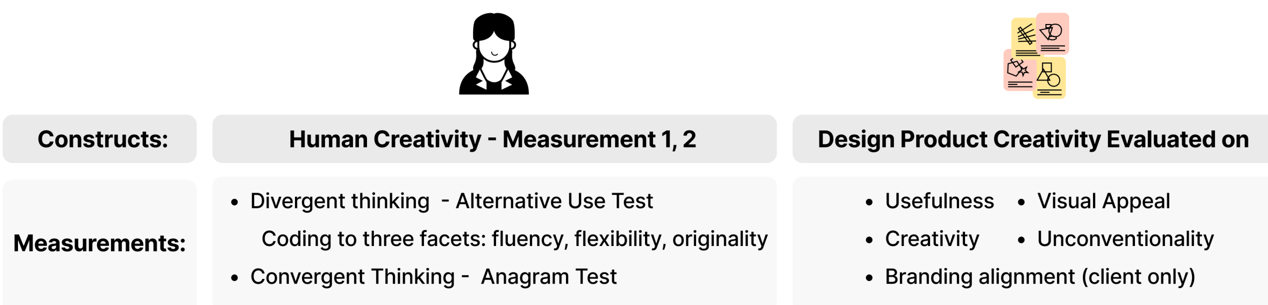

Figure 2: We operationalized creativity into two components: human inherent creativity (Study One) and design product creativity (Study Two). Human inherent creativity was measured twice per participant: once after designing with Generative AI (GenAI) and once without (human-only). This was assessed using the Alternative Uses Test [4] (for divergent thinking, coded into three facets: fluency, flexibility, and originality) and the Anagram Test [12] (for convergent thinking). In Study Two, design products collected from Study One were evaluated by five client representatives and 155 online volunteers, who rated each design's usefulness, creativity, unconventionality, and visual appeal. Clients also evaluated brand alignment with their organization.

solution [11–13, 60]. Frequently convergent thinking is assessed through Anagram Test [40, 60], where participants rearrange letters to form words, for example, "arc" translates to "car". Both AUT test and anagram tests have been used to understand people's creativity change in interventions [20, 40, 60, 66, 72, 104], showing people's creativity are sensitive to temporary modifiable factors.

With respect to the *Product* aspect, creativity is usually evaluated by the novelty and usefulness of the outputs, often externally by reviewers rather than self-assessed by creators [23, 25]. In visual design and marketing contexts, prior research also assesses design products' alignment with client branding [26, 27, 37] and overall visual appeal [37, 82]. For example, Dow et al. compared the effectiveness of advertisements created via serial prototyping (five prototypes, each followed by feedback) to the effectiveness of advertisements created via parallel prototyping (three prototypes then feedback, two more then feedback, then a final design output). Using real-world click-through data, expert ratings, crowdsourced ratings, they found that parallel prototyping yielded higher-performing ads than serial prototyping [27].

In addition, instead of treating creativity as a static trait and only judging final products, Kupers et al. developed a micro-coding method that segments a creative activity into small, meaningful units (e.g., utterances or actions) and codes each unit on two ordinal scales: a novelty scale (from repetition to entirely new ideas) and an appropriateness scale (from off-task to fully aligned with task constraints). They show that the coded units can be analyzed using time-series plots and state-space grids to reveal patterns, transitions, and variability in real-time creative activities. The approach provides a flexible, domain-general framework for quantifying creativity "in the here and now" [56]. Furthermore, Mahler proposes a domain-independent formalized metric for evaluating creativity that can be applied to artifacts produced by humans, computers, or human–computer collectives. The formalization focusing on three necessary qualities of any creative artifact: novelty, value, and unexpectedness. The goal is to shift creativity research from task-specific judgments to a shared, generalizable metric for assessing creative products [65]. Recently, several studies have begun to probe GenAI's role in creativity [8, 17, 25]. For example, using a 14-item expert-designed Torrance Test for Creative Writing, one paper shows that human-written stories pass 85% of creativity criteria, whereas LLM stories pass only 9–30%. This indicates that LLMs are less likely to meet expert creativity standards and are not reliably creative [17].

Drawing from empirical literature and established measurements, our study uses the AUT and anagram tests to assess how GenAI tool usage influences participants' creativity (see section 3 for details). For evaluating design output, we recruited clients and online volunteers who were blind to the experiment conditions to rate the designs based on its usefulness, creativity, unconventionality, visual appeal, and client' brand alignment (see section 5 for details).

2.2 Creativity Support Tools with Generative AI

With the advancement of computer technology, modern software tools have been created to support creative processes, which have been classified broadly as *creativity support tools* (CSTs) [30, 41, 90, 91]. CSTs are designed to support both creativity in individuals and collaboration in teams, aiming to *"empower users to not only be more productive, but more innovative"* [84]. Today, CSTs are widely used in digital design, with tools like Adobe's Creative Suite (Photoshop, Illustrator, etc.) [59], Canva [16] as the most popular.

Recently, more and more CSTs, especially these popular ones, are rapidly incorporating generative AI features to assist designers [1, 2, 15, 29]. They promise to provide support across design stages, from idea exploration [44] to prototyping [49]. In marketing and advertisement design, studies have found that GenAI-created imagery has potential for better quality designs [37] and also the potential to transform the industry into a data-driven, tool-based, synchronized, and more efficient system [77]. However, research also shows that perceived eeriness in AI visuals negatively impacts consumer acceptance, especially if images of people look almost, but not fully, human[34].

At the same time, a growing body of work cautions that integrating GenAI into existing design workflows is far from straightforward. In a study investigating why even experienced CAD designers struggle to co-create with GenAI design tools, Gmeiner et al. found that designers have difficulty interpreting AI outputs, lack clarity about how their inputs shape results, and often resort to workarounds or abandon advanced features because the tools provide little scaffolding for learning or shared decision-making [33]. In a within-subjects study, Kulkarni et al. asked 14 pairs of employees to design party invitations once with image search only and once with a GenAI diffusion model. By analyzing participants' interactions, artifacts, and post-task reflections, the authors found that GenAI supports broader design-space exploration and fluid remixing of collaborators' ideas but introduces challenges around fine-grained control, asymmetric access to prompts, and nondeterministic outputs [54].

Despite this surge of work on new GenAI powered CSTs, there remains a scarcity of prior work validating their influence on both the novelty and usefulness of design outputs.

2.3 Designers' Perceptions and Concerns about GenAI Tools

Recent studies have explored the opportunities and challenges of GenAI in design. Interviews with visual artists find that artists value GenAI for automating the creation process, reducing repetitive tasks, supporting the exploration of creative ideas, and facilitating communication with multiple stakeholders [51]. A study with UI/UX designers revealed that designers highly value the human factors of "enjoyment" and "agency" when working with GenAI [57]. Researchers have proposed that AI-supported design systems incorporate features such as inspiration search, exploration of design alternatives, customization of design systems, and automated checks for design guideline adherence [62]. GenAI tools are particularly effective during the divergent phases of design, helping designers overcome creative blocks and explore new possibilities. However, designers feel they are less suited for convergent tasks that require precise iteration [98].

One promising approach to human-AI collaboration that researchers have noted is the concept of mixed-initiative co-creativity [23, 106]. In this paradigm, human and AI agents work together, each contributing according to their strengths. Studies have investigated how mixed-initiative systems can support various aspects of the creative process, such as idea generation, conceptual exploration, and design refinement [22, 52, 58].

However, the impact of generative AI on mixed-initiative creativity is still a highly contested topic of ongoing research. One study found that access to GenAI worsened creative performance for participants whose creative potential was above the sample median [111]. While some researchers suggested that generative AI can enhance individual creativity by providing inspiration and facilitating exploration through suggesting novel and unconventional ideas [39], others found that GenAI outputs can be overly generic [57] and fear that GenAI may homogenize creative outputs [99], and reduce conceptual diversity at a collective level [25],

Additionally, others caution that an overreliance on AI might constrain human creativity by promoting conformity or limiting exploration [43, 46], and by encouraging cognitive offloading and dependence [32]. Junior designers also worry about ethical implications, skill degradation, limited opportunities to develop core design skills, and the potential for being replaced by GenAI tools [57, 98].

Moreover, most existing studies on AI's impact on creativity have focused on textual outputs using text-to-text models, like writing essays [74], fictional stories [25], or imaginative and experimental scenarios [8, 9]. However, there still remains a notable gap in research applying AI to real-world text-to-image design tasks and evaluating these outputs to assess their creativity.

Our study addresses this limitation by examining text-to-image generation in real-world design contexts, collecting data on both designer experiences and creative outputs to provide an understanding of GenAI's impact on both designers and their design outputs.

3 Study One: Experimental Method

In Study One, we conducted a within-subjects experimental study, in which 36 participants designed social media advertisements to promote a newly launched website for a research group (see an example output in Figure 3, and other examples in Figure 7 in the Appendix). All participants participated in each of two conditions: one in which they designed social media advertisements for the research group *with* the support of GenAI, and one in which they designed social media advertisements *without* it. In the GenAI-supported condition, participants were instructed to use AI to assist their design process. As detailed below, in both conditions participants were allowed to use additional non-AI tools to refine their designs as needed (e.g., adjusting text, layout, or other visual elements). All participants completed both sessions, with session order counterbalanced.

3.1 Participants

We recruited 36 participants via academic communication channels (Slack, Discord, etc.), email lists, and by posting fliers at a large public university campus in the United States. We recruited participants primarily from a college community and alumni networks; we did not specifically target professional designers. We advertised the study as a creativity and design study and included a QR code on the fliers and online posts that directed interested individuals to an online screening survey. 170 potential participants filled out the screener. In this initial survey, participants self-reported their design experience, typical design output types, prior experience using AI for design work, gender, native language, current or completed college major, and other demographic information. For design experience, they rated themselves on a 1–5 scale (from "minimal" to "expert"). We included participants who reported prior experience with graphic design tasks (e.g., posters, website design), and we aimed to recruit participants with varying levels of design experience and prior exposure to AI tools. All participants included in the study had some previous graphic design experience. Participants were over-representative of young people, with 94% ($N = 34$) under the age of 30, and of women (81%, $N = 29$). They had a diverse range of AI design tool experience with 67% ($N = 24$) having at least some experience using GenAI for design tasks. Each participant received a $40 Amazon gift card for completing the study. Participants' self-evaluated design ability and other design background information are shown in Table 2 (Appendix). We include additional participant demographics in Table 3 (Appendix).

3.2 Apparatus

The design tasks were conducted via an in-lab-controlled environment. Participants in GenAI-supported condition were asked to use the ChatGPT web app [70], powered by the DALL-E3 text-to-image model on their laptop, to generate and modify designs (we provided participants with ChatGPT Plus accounts). The decision to control the GenAI tool (e.g. ChatGPT) was driven by two practical considerations: the need to purchase subscriptions for ChatGPT Plus before the experiment and the need for participants to have baseline familiarity with the tool (ChatGPT text-to-image is both simple to use and has gained widespread popularity). For the human-only session, participants completed their designs individually using their preferred software tools, reflecting their realistic design workflows. During the human-only condition session, we observed participants typically used Adobe Photoshop, Figma, Canva, and for a minority used Adobe Illustrator and Google Slides; participants could also search online for existing design assets (e.g., stock icons, images) as needed. Participants completed all assessments using Qualtrics, an online survey platform [78].

We performed assessments in two areas: evaluations of participants' intrinsic creativity (divergent and convergent thinking), and self-report measures of participants' experience (mood, self-efficacy).

- ***Participant Creativity:*** To measure participants' inherent creativity, we used the Alternative Uses Test (AUT) [4], an established measure of divergent thinking, and the Anagram Test [12], an established measure of convergent thinking (see section 2.1 for detailed description). In the AUT, participants brainstorm unconventional uses for a commonplace item, such as a bottle or table, with a two-minute time limit. This is then repeated with a second item. The total number of uses participants list across both objects is their final score. Participants received the following instructions:

  "*You will be given two everyday objects. Your task is to think of as many different uses for this object as possible beyond its common or intended use. Try to come up with original, unusual, diverse, and creative uses in your responses as you can. For example, if the object given to you is a paperclip, you might think of using it to: 1) eject the SIM card from a phone, 2) be material for a tiny sculpture, 3) clean small crevices. Remember, the more creative and unusual, the better!*"

  The Anagram Test assessed participants' convergent thinking by asking them to rearrange letters of a prompt word (e.g., "lamp") into a different word (e.g., "palm"). They had three minutes to rearrange as many prompt words as possible into their anagrams, with their total number solved determining their final convergent thinking score.

- ***Participant Experience:*** Consistent with prior research [27, 48], we measured participant mood via a short, self-report survey with three subscales: relaxation, pleasure, and energy level. We also measured participant self-efficacy through a self-report scale developed by Dow et al. [27].

3.3 Procedures

Study One consisted of two sequential design sessions and lasted approximately 90 minutes in total (see Study One overview Fig. 1). At the start of the session, we informed the participants that they would create advertisements intended to be used in actual advertising campaigns on social media platforms (e.g., Facebook and Instagram) to promote a research group's newly launched website. To encourage high-quality design work, we informed them that the creators of the top three designs would receive an additional \$40 gift card bonus. Participants then completed each of two study sessions:

- ***Session One:*** Participants were randomly assigned to begin either with the GenAI-supported condition or the human-only condition. In both conditions, they were tasked with designing a social media advertisement to promote a newly launched research group website communicating opportunities for the general public to participate in research. Through a short Q&A session, we showed participants example social media advertisements on a different topic and ensured they understood the design requirements. Each participant received a design task sheet with a short paragraph describing the research lab, including its history and mission, and a link to the new website. In addition, the design task sheet included the following prompt:
  - *Task.* You are asked to create a social media campaign ad for the client organization. The purpose of the campaign is to promote the newly designed website for Digital Youth Lab at University of Washington, showcasing the lab's past and ongoing research, and offering both students and parents opportunities to participate in research. You are requested to submit one or two designs (advertisements to be deployed on social media) in 25 minutes. Please make the design appropriate for a social media advertising campaign.
  - *Dimensions.* The advertisement should be formatted to a size of 1080×1080 pixels (1:1 aspect ratio).
  - *Mandatory text.* Each advertisement must include the following text: 1) "Digital Youth Lab" 2) "Explore new studies and participation opportunities," and 3) an institution logo.

  In the GenAI-supported condition, participants were instructed to use ChatGPT for design support, but were also permitted to use other tools to correct inaccuracies in AI-generated text, clip images as needed. Participants in the human-only condition completed their designs without any AI support. Each participant had 25 minutes to create one or two designs, a duration determined to be feasible through prior pilot testing. After completing the design task, participants immediately completed self-report measures assessing their mood and self-efficacy (typically taking under 30 seconds). They then completed the Alternate Uses Test (AUT) to assess divergent thinking, followed by the Anagram Test to assess convergent thinking. All assessments were administered via Qualtrics. After these assessments, participants took a 1–3 minute break before starting Session Two.

- ***Session Two:*** This phase mirrored the first, with participants completing the condition they did not complete in Session One: those who used ChatGPT initially (GenAI-supported condition) now designed individually (human-only condition) and vice versa. After completing the second design session, participants were once again evaluated on their mood, self-efficacy, divergent thinking via the AUT, and their convergent thinking via the Anagram Test. We collected all design outputs from both sessions for subsequent evaluation in Study Two. When we collected their designs, participants were also asked to indicate which of their designed advertisements was their "best" design (only one per participant).

3.4 Data Analysis

We conducted within-subjects analyses with one primary factor: design support (GenAI-supported vs. human-only). We also examined two participant subgroups: prior GenAI tool experience (with prior GenAI design tool exposure vs. without prior GenAI design tool exposure) and native language proficiency (native vs. non-native English speakers). We examined GenAI's impact on participants' creativity and experience both across the entire participant group and within these subgroups.

During the analysis, consistent with the literature, one researcher coded participants' AUT responses for three dimensions: fluency, flexibility, and originality using the counting method in [4]. To evaluate differences between experimental conditions, we performed Wilcoxon signed-rank tests on scores of divergent thinking (AUT), convergent thinking (Anagram Test), mood (relaxation, pleasure, energy), and self-efficacy.

4 Study One Results: Participants' Experience

4.1 Overall Effects on Participants' Mood, Self-Efficacy, and Creativity

We found that there were no significant differences between conditions in mood (relaxation, pleasure, energy), self-efficacy, inherent divergent thinking (AUT), or convergent thinking (Anagram). However, further subgroup analyses revealed significant condition differences based on prior experience with GenAI tools and native language proficiency.

4.2 Subgroup Analysis: Prior Experience with GenAI

A Wilcoxon signed-rank test showed that, for participants without prior GenAI design tool experience, the 25-minute GenAI-supported design session significantly enhanced divergent thinking. Our analysis showed a significant increase in idea fluency score ($M = 14.333$, $SD = 4.384$) and flexibility score ($M = 12.250$, $SD = 3.562$) in the GenAI-supported design session, as compared to the human-only design session (fluency, $M = 12.500$, $SD = 4.573$, $Z = -2.112, p < .05$; flexibility, $M = 10.167$, $SD = 3.131$, $Z = -2.203$,

Figure 3: Example social media advertisement designs created by participant 8. The left design was produced without AI support; the right design was produced with AI support. In Study Two, clients and online volunteers evaluated the designs based on their usefulness, creativity, unconventionality ("weirdness"), alignment with the client's brand (only evaluated by the clients), and visual appeal. See additional design examples in Figure7 in the Appendix.

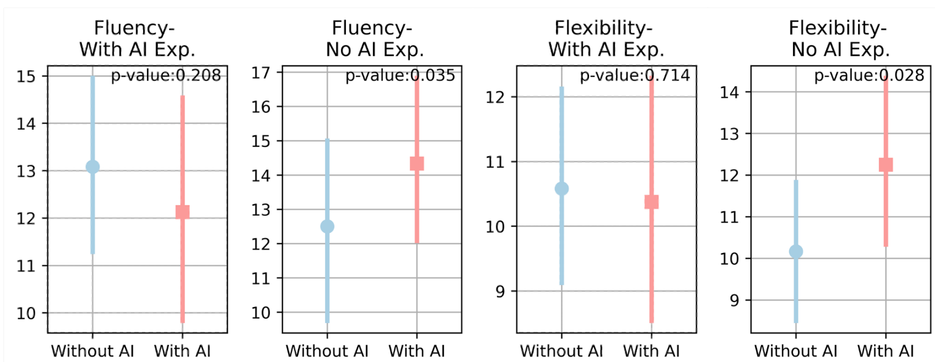

Figure 4: Analysis of the impact of GenAI on divergent thinking measured by idea fluency and flexibility for subgroups with and without prior GenAI experience. For participants with no prior GenAI design tool experience, GenAI significantly increased idea fluency ($p < .05$) and flexibility ($p < .05$) compared to the human-only condition. No significant improvement in fluency or flexibility was observed for participants with prior GenAI experience. Error bars represent 95% confidence intervals.

$p < .05$). However, for participants with prior GenAI design tool experience, we observed no significant differences between conditions in divergent thinking scores. Additionally, there were no significant condition differences in convergent thinking, mood, or self-efficacy regardless of prior GenAI experience.

5 Study Two: Design Evaluation Method

5.1 Participants

To understand GenAI's influence on design output, we recruited two groups: (1) evaluators who were members of the research group whose newly launched website participants aimed to advertise ("*Clients*"). And (2) online volunteer participants to evaluate the advertisement design products created during Study One.

Clients: To evaluate the designs produced by participants in Study One, we recruited five members of the research group that the advertising campaign aimed to promote, serving as "clients" and stakeholders for this design work. We recruited evaluators by posting recruitment messages to the client lab's communication channels and through word-of-mouth. We recruited client evaluators to enhance ecological validity and to provide specific perspectives on design evaluation such as their evaluation of brand alignment with the research group. Evaluators were researchers actively involved in the research group for which the advertisements were created, and all had backgrounds in design and technology. They were stakeholders who could determine which advertisement to deploy. Evaluators also provided qualitative feedback about the design products, as illustrated in Section 5.3. Evaluators received a $40 gift card as a thank-you for their time. Demographic information about client evaluators is shown in Table 4 in the Appendix.

Online Volunteers: We also recruited online volunteers through the online research study platform LabintheWild[1] Consistent with

[1]https://labinthewild.org/ The research platform has previously been used for design evaluations [36, 79–81], with the data quality being comparable to controlled lab studies and superior to those conducted on Mechanical Turk [81, 107].

standard practice on this research platform, at the end of the evaluation, we provided each participant a comparison visual of their ratings to the average ratings previously collected from client evaluators. This feedback served as an incentive and acknowledgment of their participation on this research platform. In total, we collected ratings from 155 volunteer participants after excluding incomplete entries. Participants were aged 9 to 64 years ($M = 25.6, SD = 11.5$)[2]. The majority were from the United States ($N = 62, 40.0\%$), with the remainder from other countries ($N = 93, 60.0\%$). The gender distribution was: female n=90 (58.1%), male n=56 (36.1%), non-binary n=3 (1.9%), prefer not to say n=5 (3.2%); one participant selected other (0.6%).

5.2 Apparatus and Procedure

Both client evaluators and online volunteers completed their evaluations online.

- ***Design Product Evaluation by Clients:*** To evaluate the quality of designs produced by participants in Study One, we created a Likert scale survey for each design product to be completed by five client evaluators blind to experiment conditions. The survey included an 11-point scale question about the usefulness of the design ("*In a social media ad campaign, how well will this ad perform?*"), and four 7-point "strongly disagree" to "strongly agree questions" to measure designs' creativity ("*This design is creative*"), unconventionality ("*This design is weird*"), brand alignment with clients' organization ("*This design effectively represents our organization,*") and visual appeal ("*This design is visually appealing*"). In addition, the survey included three free-response textboxes per design where clients could optionally input their qualitative comments such as likes, dislikes, and explanations for their ratings. Each client was presented with 105 designs in random order using Qualtrics. Evaluating all designs took approximately 45-75 minutes, with breaks allowed as needed.
- ***Design Product Evaluation by Online Volunteers:*** Similarly, we created a Likert scale survey for each design produced by Study One. The survey included four 7-point "strongly agree" to "strongly disagree" questions to evaluate each design's usefulness ("*If I saw this advertisement on social media, I would click it*"), creativity ("*This advertisement's design is creative*"), unconventionality ("*The advertisement looks weird*"), and visual appeal ("*This advertisement is visually appealing*"). We deployed the study using the online volunteer sourcing research platform LabintheWild from April to August 2025. Each volunteer participant completed one practice rating before evaluating ten randomly presented designs. Only ratings from these ten designs were included in our final analysis.

5.3 Analysis

To analyze both clients' and online volunteers' perceptions of design outputs, we used a Generalized Linear Mixed Models (GLMM). Our measured dependent variables (creativity, unconventionality, visual appeal, usefulness, and brand alignment) were collected using ordered Likert scales, and the data exhibit hierarchical non-independence owing to 1) each rater rated multiple designs, and 2) each participant could submit multiple designs, resulting in designs being nested within participants. A GLMM is appropriate and can simultaneously model fixed effects (presence of GenAI support) and random effects (individual differences among raters, and variability across designs nested within participants). Our model can be represented as: Outcome ∼ GenAI_support + (1 | rater) + (1 | design participant) + (1 | design participant:design).

To analyze the qualitative open-ended feedback provided by client evaluators, one researcher performed a thematic analysis [14] on all qualitative comments from the five client evaluators. The researcher began by carefully reviewing each comment and adding detailed notes to capture initial observations and ideas. After this initial open-coding phase, the researcher iteratively refined the identified codes, grouping similar comments and merging related codes into broader, coherent themes.

Table 1: Effect of GenAI support on design evaluations (GLMM results). Significant effects bolded.

	Clients				Volunteers			
	β	SE	z	p	β	SE	z	p
Creativity	.81	.18	4.41	**<.001**	.90	.15	6.04	**<.001**
Unconventional	1.30	.19	6.72	**<.001**	.66	.11	6.22	**<.001**
Visual appeal	.26	.18	1.43	.153	.45	.14	3.25	**<.01**
Usefulness	.02	.16	0.12	.905	.00	.11	0.01	.995
Brand align.	-.19	.20	-0.92	.359	—	—	—	—

6 Study Two Results: Design Product Evaluation

In total, participants generated 110 designs (54 designs in the human-only condition and 56 designs in the GenAI-supported condition). Five designs were filtered out for non-compliance with the study requirements. This resulted in a final count of 105 eligible designs (52 human-only and 53 GenAI-supported) for design evaluation. When submitting their work, participants were also asked to indicate which of their own advertisements they considered their "best" design. They selected a GenAI-supported design in 17 of 36 cases (47.2%), with the remaining 19 preferring a human-only design. This near-even split suggests that, from the participants' own perspective, GenAI support did not consistently produce designs they viewed as better. The average number of designs per participant was 1.44 in the human-only condition and 1.47 in the GenAI-supported condition. Five client evaluators each rated 105 designs, yielding a total of $N = 525$ ratings. Additionally 155 online volunteers each rated 10 designs, yielding a total $N = 1550$ ratings.

We first present results from the client evaluations, followed by evaluations from the online volunteers. Detailed comparisons between client and volunteer evaluations are presented in Table 1.

[2]To enable anyone to participate, including minors, the research team obtained a waiver for parental consent from their IRB.

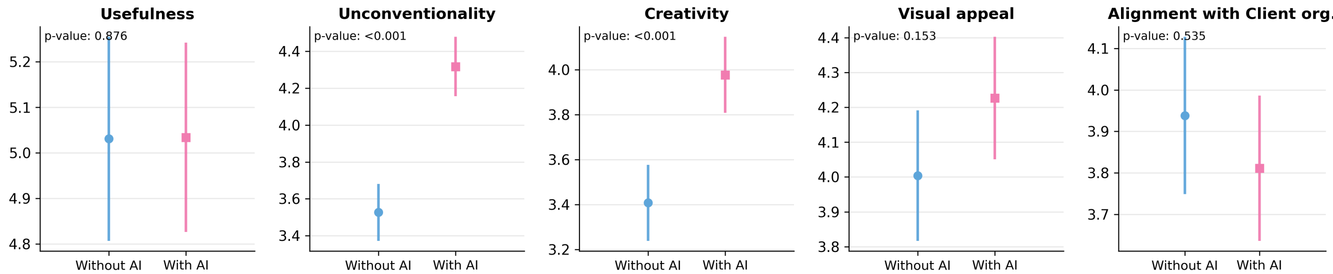

Figure 5: Client Evaluation: Analysis of design output evaluated by clients in GenAI-supported vs. human-only conditions. There is no significant difference in the perceived usefulness, visual appeal, and brand alignment between the conditions. However, designs supported by GenAI were rated significantly more creative ($p < .001$) and unconventional ($p < .001$). Error bars represent 95% confidence intervals.

6.1 Client Evaluation: Creativity, Unconventionality, Usefulness, Visual Appeal, and Brand Alignment

We used a GLMM to analyze how GenAI support influenced client perceptions of the creativity, unconventionality, usefulness, visual appeal, and alignment with the client's organization for the 105 designs. The models included random intercepts for raters ($n = 5$) and designs nested within participants ($n = 105$ designs from $n = 36$ participants), addressing non-independence among designs by the same participant. Client raters perceived GenAI-supported designs as significantly more creative ($\beta = 0.805$, $SE = 0.183$, $Z = 4.405$, $p < .001$) and more unconventional ($\beta = 1.300$, $SE = 0.193$, $Z = 6.721$, $p < .001$) than human-only designs. However, clients reported no significant differences between GenAI-supported and human-only conditions in perceived usefulness ($\beta = 0.019$, $SE = 0.163$, $Z = 0.119$, $p = .905$), visual appeal ($\beta = 0.264$, $SE = 0.184$, $Z = 1.430$, $p = .153$), or alignment with the client's organization ($\beta = -0.188$, $SE = 0.204$, $Z = -0.917$, $p = .359$). See Fig. 5 for visualization.

6.2 Online Volunteer Evaluation: Creativity, Unconventionality, Usefulness, and Visual Appeal

We applied the same modeling to analyze online volunteers perception of creativity, unconventionality, usefulness, and visual appeal for the same 105 designs (note alignment with client measure does not apply for online volunteers). Again, random intercepts account for variability across individual raters ($n = 155$) and designs nested within participants ($n = 105$ designs from $n = 36$ participants). Online volunteers rated GenAI-supported designs as significantly more creative ($\beta = 0.900$, $SE = 0.149$, $Z = 6.040$, $p < .001$), more unconventional ($\beta = 0.658$, $SE = 0.106$, $Z = 6.218$, $p < .001$), and more visually appealing ($\beta = 0.447$, $SE = 0.138$, $Z = 3.245$, $p < .01$), compared to human-only designs. However, there was no significant difference between GenAI-supported and human-only conditions in perceived usefulness ($\beta = 0.001$, $SE = 0.112$, $Z = 0.006$, $p = .995$). See Fig. 6 for visualization.

In summary, the findings show both clients and online volunteers perceived GenAI-supported design as more creative and unconventional, suggesting a clear novelty effect by involving GenAI. However, neither group observed improvements in perceived usefulness of the designs, indicating there is a decoupling between design perceived novelty and usefulness.

6.3 Qualitative Analysis of Client Evaluation Comments

6.3.1 Perceptions of GenAI-Supported Designs: Increased Creativity and Colorfulness with Potential Drawbacks of Complexity and Confusion. Clients generally found GenAI-supported designs to be more creative and engaging than human-only designs, frequently praising their "*very creative and appealing*" style (E1). The "*bright colors*" (E3) and "*eye-catching compositions*" (E5) were highlighted as key elements that made these designs stand out. They mentioned these designs potentially could attract viewers in contexts like social media advertisements, as one client noted, "*I'd guess imagery is much more noticeable as a brief ad*" (E1). The use of vibrant colors was consistently seen as a strength, with descriptions such as "*visually interesting, attractive,*" (E1) and featuring "*flashy colors and cute robots*" (E5) that captured attention. Even when some clients were not enthusiastic about certain designs overall, they still recognized that the colors were "catchy" and visually compelling (E3). Client evaluators attribute these effects largely to GenAI's innovative use of colors and layouts, which were perceived as more dynamic and engaging (E1, E3).

However, GenAI-supported designs sometimes led to issues of being too complex and lack of clarity. Clients criticized several GenAI-supported designs for being "*too busy*" (E3, E5) and having "*too many elements*" (E1) that cluttered the visual space, making them difficult to interpret quickly. Comments noted that there were "*too many colors and icons to look at*" (E1), creating a sense of visual overload. Additionally, text legibility was a recurring concern; clients found that the text was often "*not legible,*" which further detracted from the overall effectiveness of these designs (E1).

While GenAI-supported designs mostly followed graphic design norms, they sometimes broke them through unnatural or abnormal compositions. For example, clients were puzzled by choices like images "*overlaying another image that is obstructed*" (E1), noted that appeared "*2D but the actual ad is oriented away from the viewing angle*" (E2), and background elements such as graphs that did not

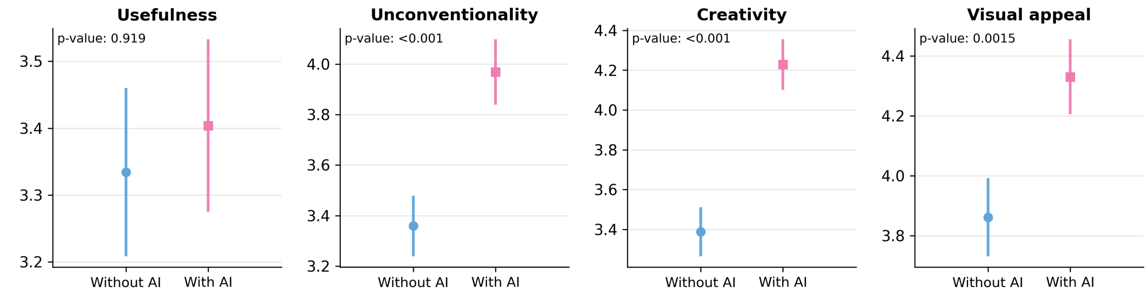

Figure 6: Online Volunteer Evaluation: Analysis of design output evaluated by online volunteers in GenAI-supported vs. human-only conditions. There is no significant difference in the perceived usefulness. However, designs supported by GenAI were rated significantly more creative ($p < .001$), unconventional ($p < .001$), and visually appealing ($p < .01$). Error bars represent 95% confidence intervals.

seem to add meaningful content to the poster (E3). Moreover, certain designs were described as "*weird*" or "*uncanny,*" with unusual compositions or elements that felt out of place or unsettling, as one client said, "*the face looks a bit too uncanny valley for my taste*", and also mentioned another design had "*a weird artifact on the image by [the text] 'innovation'*" (E5).

These aspects corroborate our quantitative findings suggesting that while GenAI-supported designs are perceived as creative and visually striking, they sometimes fail to maintain clarity and are perceived as too busy, potentially limiting their usefulness in communicating intended messages.

6.3.2 GenAI-supported Design Broadens Inclusivity, But Sometimes Misunderstands Social Norms. Clients noted that GenAI-supported designs effectively showcased diversity, featuring a wide range of technologies and representations of children from various backgrounds. One client appreciated the "*images showing the diversity of technologies and multiple children*" (E1), and another mentioned, "*representation of diverse teens*" (E3), highlighting the broader inclusivity in these designs. In contrast, some human-only designs relied on more stereotypical imagery, such as the "*overuse of stock imagery of white human hands with white robot hands*" (E3). These observations suggest that GenAI has the potential to expand representation and diversity in these design tasks.

However, clients also pointed out that some GenAI-supported designs misaligned with social norms and could unintentionally convey inappropriate messages. For example, one client found certain characters problematic, stating that people next to children "*look like adults,*" and if so, "*the hand on the child is immediately disturbing*" (E1). Another comment noted that certain images appeared "*super AI generated, and not in a good way,*" questioning "*why are the teens hugging? And what's in the background?*" (E3). These critiques show that while GenAI designs introduce greater diversity, they can also create uncomfortable or ambiguous imagery, potentially leading to unintended interpretations.

7 Discussion

7.1 GenAI's Impact on Decoupling Novelty vs. Usefulness of Design Outputs

Creativity fundamentally requires both novelty and usefulness [61, 63, 73, 88]. Both client and volunteer evaluators in our study rated GenAI-supported advertisements as significantly more creative and unconventional compared to human-only designs. However, neither group judged them more useful, and clients noted no brand-alignment gain. This divergence highlights an important and surprising problem: a decoupling between novelty and usefulness in GenAI-supported visual design.

Several factors may explain this decoupling effect. Clients' qualitative comments frequently described GenAI-supported designs as visually appealing, yet often criticized them as "*too busy,*" or having "*too many colors.*" Clients' feedback shows a tension between aesthetic appeal and clarity, as eye-catching visuals, clutter, and unconventional composition can sometimes overshadow a design's message, diminishing their usefulness. Additionally, some outputs generated by GenAI displayed ambiguous or uncanny elements that violated social norms, further detracting from their perceived appropriateness. Another likely factor is the hidden optimization of current commercialized GenAI models for visual interest. Commercialized models tend to share certain aesthetics [68], prioritize surprising, attention-grabbing visuals rather than clearly supporting design goals (e.g. Midjourney founder mentions their model's style is "a bit whimsical and abstract and weird...surprising and beautiful." See [100]), leading to outputs that may be intriguing but fail to consider practical design requirements and objectives.

Our findings also highlight several potential downsides of widespread GenAI use in design. Prior work describes the rise of "AI slop," where low-quality, machine-generated content saturates online platforms [64]. Our results suggest that graphic design is one domain where this risk is very real: GenAI can produce novel ads that are not, on average, more useful. Qualitatively, we observed "weird" and sometimes socially inappropriate visuals that could pollute the contemporary design space and, over time, shape emerging users' visual taste and broader cultural norms. We also saw recurring patterns such as very bright colors and complex compositions, hinting at a converging "AI style," which echoes concerns that generative AI homogenizes outputs [99] and reduces diversity at a collective level [25]. Designs perceived as creative now may quickly come to feel banal once this aesthetic saturates the culture.

Using GenAI in design also places metacognitive demands [94] on users, requiring sustained monitoring and control. Novelty can create an illusion of quality: designers, clients, and end-users are potentially vulnerable to a novelty bias, mistaking different for better. In our study, this sometimes led to submissions that were eye-catching but unclear, or misaligned with social norms. To counter this, designers need a separate evaluation pass that deliberately asks: Is this legible? On message? On brand? Accessible? Appropriate?—that is, engaging "System 2" thinking rather than relying on immediate impressions [45].

Our participants were primarily novice designers recruited from a college community and alumni networks. In this setting, our results complicate the common assumption that integrating GenAI into design workflows reliably improves downstream effectiveness. Our results motivate a shift toward outcome-based evaluation, such as measuring brand comprehension, clarity, and customer satisfaction. Future work should also examine longer-term learning trajectories to understand how GenAI affects the development of design judgment and effectiveness over time for novice designers. For example, GenAI tools can also serve as learning-oriented systems that provide structured feedback on clarity, brand alignment, and accessibility, rather than optimizing primarily for novelty.

More broadly, our results underscore that novelty is not inherently good; it is conditional. Novelty is beneficial only when basic constraints are respected, the message remains clear, and the audience is not overloaded or confused. If GenAI tends to overshoot on novelty, the designer's role shifts toward curating, constraining, and repairing AI outputs rather than simply accepting them. This perspective helps justify why human judgment remains central: the designer is responsible for turning "interesting" into "appropriate and useful." This is especially important for novice designers entering the field through GenAI tools as their long-term skill will depend on recognizing when a GenAI design fails and knowing how to fix it.

Finally, our findings point to an opportunity for GenAI design tools to shift from just image generators into critique partners. Instead of only producing alternatives, tools could analyze a design task and the resulting designs, flagging issues such as low contrast, illegible text, platform-inappropriate formats, or weak brand alignment, and offering explanations of these trade-offs. AI research suggests that models are often better at critiquing than creating on the first attempt [103], which makes this direction especially promising. As more novice designers rely on GenAI, having tools that also teach what is appropriate and effective, rather than only what is possible, will be critical.

7.2 Human Creativity in the Age of GenAI

We observed heterogeneous effects in influencing participants. Participants new to GenAI demonstrated higher divergent thinking after a 25-min GenAI design session, specifically in fluency and flexibility. Conversely, those with prior AI experience showed no comparable short-term improvement. This suggests that generative AI's cognitive impact varies based on users' prior experience. While our findings support optimism of GenAI adoption for participants who have never used GenAI-powered design tools, this advantage appears temporary, diminishing as participants become accustomed to the tools. This observation aligns with previous research indicating that LLMs may improve short-term creative performance but potentially impair independent creative ability over time [55]. These findings imply that the creative benefits of GenAI may not persist after users gain familiarity with GenAI tools, and reliance might even negatively impact users' creativity in the long run.

In addition, building on previous literature, our results indicate that GenAI might induce false confidence in users. Although our participants did not report improved self-efficacy following short-term GenAI usage, various prior studies report designers often perceive enhanced confidence and reduced anxiety and cognitive load when using AI tools [3, 42], particularly in early-stage, concept-development tasks [18]. Such perceived ease of production, however, risks inflating users' confidence in their creative abilities without improving actual creative skills or outcomes [31]. Our results signal that this might be true, since the use of AI does not lead to a perceived increase in design usefulness. This "illusion of creative competence" may be especially problematic in educational contexts. If novice designers misattribute AI-driven improvements to their own skill, they risk underdeveloping creative abilities such as design judgment, critical evaluation, and aesthetic taste. Furthermore, as users increasingly adopt project-level orchestration roles (e.g., project manager) when working with generative AI, they encounter challenges in clearly defining and articulating creative intent and goals [75], which is critical for the design to be useful. These concerns raise questions about how users working with GenAI tools can develop and grow their creative skills rather than relying on temporary, potentially misleading boosts in self-confidence.

The emergence of an AI-only workflow further complicates discussions surrounding human-AI co-creation with GenAI tools. A recent marketing research study shows that advertisements generated by AI alone (e.g., users only implement the workflow) outperform advertisements generated by humans alone on measures such as perceived quality, realism, and advertising effectiveness. This field study shows that banner ads designed by AI have higher click-through rates than human-made banner ads [37]. Our findings introduce additional complexity to this discourse, showing that although GenAI-supported designs displayed higher perceived novelty than human-only designs, they did not improve perceived usefulness. With prior literature, our results challenge prevailing assumptions about human-AI co-creation as an inherently beneficial paradigm in the CST domain. Future research should therefore clarify the contexts, use-cases, and conditions under which AI-only, human-only, or human-AI co-creation processes are optimal. Future empirical work should also explore the unique value of human creativity, such as taste, intuition, and cultural sensitivity that may resist or complement the characteristics of AI-generated output.

At a societal level, the widespread adoption of GenAI in creative industries raises concerns regarding long-term diversity in creative outputs and public acceptance. Literature shows that GenAI can homogenize the ideas generated by different users [8] and limit human agency and lead to creative fixation [5, 21, 101]. Reliance on a few dominant generative models such as Nano Banana or Midjourney may inadvertently homogenize design outcomes [25], limiting stylistic diversity and narrowing the exploration of design solution space. As AI-generated visuals proliferate, initial perceptions of novelty may diminish, potentially leading to aesthetic fatigue among audiences. This raises critical questions: Is the current hype of GenAI-generated designs merely a temporary novelty effect or will it persist over time? Could sustained exposure lead viewers to value human-only creations more, driven by fatigue with AI aesthetics or a desire for authenticity? These questions call for longitudinal research to examine how acceptance evolves as GenAI content continues to saturate public media and cultural spaces.

Creativity holds intrinsic value, allowing people to express unique human perspectives, emotions, and experiences [71]. Through creative endeavors like writing, art, music, and problem-solving, individuals can explore and articulate who they are [87] alongside providing a sense of authenticity and fulfillment [47, 92]. As generative AI increasingly mediates creative tasks, it becomes essential to critically evaluate how these technologies impact, support, or potentially constrain our innate creative capacities on both individual and societal levels. As Kaufman once said:

> "*The way we define and study creativity has deep implications for how we see ourselves—as more or less agentic beings, as determined by our society and culture or actively shaping it, as different from or similar to the divine*" [47].

8 Limitation and Future Work

One limitation of our study is the reliance on subjective evaluations by client and online volunteer raters to assess the perceived usefulness of the design products. Although subjective ratings are common proxies for real-world design effectiveness [26, 50], they may not fully capture objective performance metrics, such as actual click-through rates on social media platforms. Nevertheless, the strong alignment between client and online volunteer ratings in our study provides converging evidence. Future studies should further validate these subjective evaluations by examining objective real-world outcomes, such as user engagement or behavioral metrics.

Second limitation was our use of the GPT-4 model powered by DALL-E3 for text-to-image generation, rather than other CSTs powered by models specifically designed for design tasks. We selected ChatGPT due to its popularity, widespread familiarity, and accessibility. This limitation reduces our generalizability since designers may use different kinds of creative support tools. Future research should investigate the impacts of other text-to-image or image-to-image GenAI models, dedicated visual design tools, and other general-purpose chatbot-based assistants. Third, our study is the controlled lab environment itself. Although participants in the human-only condition freely selected their preferred design tools to respect their typical workflows, the study is still in an artificial setting. Fourth, the qualitative analysis of designs was based on optional free-response text from only five clients, a limited dataset from which to derive in depth insights. Fifth, our participant sample was drawn primarily from a college community and alumni networks; while many participants had prior design experience, they were mostly design students or recent graduates. Therefore, the findings should be interpreted as evidence about novice or early-career designers, which may limit generalizability to experienced designers working in industry. Future research could recruit more experienced designers or focus on different design tasks. Future research could use other creativity metrics such as appropriateness/relatedness to evaluate designs. Additionally, future research could complement our findings by using other approaches beyond a lab study to examine designers' interactions with GenAI tools and their effects on designers and design outcomes within more naturalistic environments. Future research could also consider examining how AI disclaimers or implicit perceptions of AI-authorship shape judgments of creativity and usefulness, since knowing a design is AI-generated might influence evaluators' ratings.

9 Conclusion

As GenAI becomes more integrated into design workflows, it is crucial to understand how these tools shape both the users using these tools and the final outputs. Through a within-subjects experiment involving 36 participants, we found that while GenAI did not significantly impact the participants' overall creative thinking abilities, its influence varies among user groups with different prior AI tool experience. Participants new to GenAI tools experienced gains in divergent thinking, such as idea fluency and flexibility, while those with prior AI experience did not. We also found that designs created with GenAI support were perceived by both client and online volunteer raters as more creative and unconventional. Online volunteers also rated these designs as more visually appealing. However, neither group perceived GenAI-supported designs to be more useful, nor did clients find better alignment with brand identity compared to designs produced by humans only, suggesting a decoupling of novelty and usefulness, two key components of creativity.

These findings challenge current enthusiasm and industry trends which advocate for widespread adoption of GenAI tools in the design field. We question whether the hype of current human-AI co-creation model can actually improve users' creativity and produce better designs.

References

[1] Adobe. 2023. Next-Level Generative Fill. https://www.adobe.com/products/photoshop/generative-fill.html
[2] Adobe. 2023. Photoshop Neural Filters powered by AI. https://www.adobe.com/products/photoshop/neural-filter.html
[3] Sultan A Alharthi. 2025. Generative AI in Game Design: Enhancing Creativity or Constraining Innovation? *Journal of Intelligence* 13, 6 (2025), 60.
[4] Amin Alhashim, Megan Marshall, Tess Hartog, Rafal Jonczyk, Danielle Dickson, Janet van Hell, Gül Okudan-Kremer, and Zahed Siddique. 2020. Work in progress: Assessing creativity of alternative uses task responses: A detailed procedure. In *Paper presented at 2020 ASEE Virtual Annual Conference.*
[5] Leyla Alipour, Mohsen Faizi, Asghar Mohammad Moradi, and Gholamreza Akrami. 2018. A review of design fixation: Research directions and key factors. *International Journal of Design Creativity and Innovation* 6, 1-2 (2018), 22–35.
[6] Teresa M Amabile. 1996. *Creativity and innovation in organizations.* Vol. 5. Harvard Business School Boston.
[7] Oxford Analytica. 2023. GenAI will transform workplace tasks across industries. *Emerald Expert Briefings* oxan-db (2023).
[8] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th Conference on Creativity & Cognition.* 413–425.
[9] Joshua Ashkinaze, Julia Mendelsohn, Li Qiwei, Ceren Budak, and Eric Gilbert. 2025. How AI Ideas Affect the Creativity, Diversity, and Evolution of Human Ideas: Evidence From a Large, Dynamic Experiment. In *Proceedings of the ACM*

Collective Intelligence Conference (CI '25). Association for Computing Machinery, New York, NY, USA, 198–213. doi:10.1145/3715928.3737481

[10] Frank Barron. 1955. The disposition toward originality. *Journal of abnormal psychology* 51 3 (1955), 478–85. https://api.semanticscholar.org/CorpusID:30102436

[11] Mathias Benedek, Sabine Bergner, Tanja Könen, Andreas Fink, and Aljoscha C Neubauer. 2011. EEG alpha synchronization is related to top-down processing in convergent and divergent thinking. *Neuropsychologia* 49, 12 (2011), 3505–3511.

[12] Edward M Bowden. 1997. The effect of reportable and unreportable hints on anagram solution and the aha! experience. *Consciousness and cognition* 6, 4 (1997), 545–573.

[13] Edward M Bowden and Mark Jung-Beeman. 2003. Normative data for 144 compound remote associate problems. *Behavior research methods, instruments, & computers* 35 (2003), 634–639.

[14] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.

[15] Canva. 2023. Visualize your design ideas with MagicDesign. https://www.canva.com/magic-design/

[16] Canva. 2024. Celebrating Canva's growth in the US. https://www.canva.com/newsroom/news/canva-us-growth/

[17] Tuhin Chakrabarty, Philippe Laban, Divyansh Agarwal, Smaranda Muresan, and Chien-Sheng Wu. 2024. Art or artifice? large language models and the false promise of creativity. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–34.

[18] Tilanka Chandrasekera, Zahrasadat Hosseini, and Ubhaya Perera. 2025. Can artificial intelligence support creativity in early design processes? *International journal of architectural computing* 23, 1 (2025), 122–136.

[19] Mohammad Mohiuddin Choudhury, Boris Eisenbart, and Blair Kuys. 2025. Artificial intelligence (AI) in the design process–a review and analysis on generative AI perspectives. *Proceedings of the Design Society* 5 (2025), 631–640.

[20] Lorenza S Colzato, Ayca Szapora, and Bernhard Hommel. 2012. Meditate to create: the impact of focused-attention and open-monitoring training on convergent and divergent thinking. *Frontiers in psychology* 3 (2012), 22970.

[21] Nathan Crilly. 2019. Creativity and fixation in the real world: A literature review of case study research. *Design Studies* 64 (2019), 154–168.

[22] Nicholas M Davis, Yanna Popova, Ivan Sysoev, Chih-Pin Hsiao, Dingtian Zhang, and Brian Magerko. 2014. Building Artistic Computer Colleagues with an Enactive Model of Creativity.. In *ICCC*. 38–45.

[23] Sebastian Deterding, Jonathan Hook, Rebecca Fiebrink, Marco Gillies, Jeremy Gow, Memo Akten, Gillian Smith, Antonios Liapis, and Kate Compton. 2017. Mixed-initiative creative interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 628–635.

[24] Kees Dorst and Nigel Cross. 2001. Creativity in the design process: co-evolution of problem–solution. *Design studies* 22, 5 (2001), 425–437.

[25] Anil R Doshi and Oliver P Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* 10, 28 (2024), eadn5290.

[26] Steven Dow, Julie Fortuna, Dan Schwartz, Beth Altringer, Daniel Schwartz, and Scott Klemmer. 2011. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2807–2816.

[27] Steven P Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L Schwartz, and Scott R Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 4 (2010), 1–24.

[28] Kristen M Edwards, Brandon Man, and Faez Ahmed. 2024. Sketch2Prototype: rapid conceptual design exploration and prototyping with generative AI. *Proceedings of the Design Society* 4 (2024), 1989–1998.

[29] Figma. 2024. Your creativity, unblocked with Figma AI. https://www.figma.com/ai/

[30] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the landscape of creativity support tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–18.

[31] Yue Fu, Michele Newman, Lewis Going, Qiuzi Feng, and Jin Ha Lee. 2025. Exploring the Collaborative Co-Creation Process with AI: A Case Study in Novice Music Production. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 1298–1312.

[32] Michael Gerlich. 2025. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies* 15, 1 (2025), 6.

[33] Frederic Gmeiner, Humphrey Yang, Lining Yao, Kenneth Holstein, and Nikolas Martelaro. 2023. Exploring challenges and opportunities to support designers in learning to co-create with AI-based manufacturing design tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[34] Chenyan Gu, Shuyue Jia, Jiaying Lai, Ruli Chen, and Xinsiyu Chang. 2024. Exploring consumer acceptance of AI-generated advertisements: From the perspectives of perceived eeriness and perceived intelligence. *Journal of Theoretical and Applied Electronic Commerce Research* 19, 3 (2024), 2218–2238.

[35] Joy P Guilford. 1967. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior* 1, 1 (1967), 3–14.

[36] Lane Harrison, Katharina Reinecke, and Remco Chang. 2015. Infographic aesthetics: Designing for the first impression. In *Proceedings of the 33rd Annual ACM conference on human factors in computing systems*. 1187–1190.

[37] Jochen Hartmann, Yannick Exner, and Samuel Domdey. 2025. The power of generative marketing: Can generative AI create superhuman visual marketing content? *International Journal of Research in Marketing* 42, 1 (2025), 13–31.

[38] Andrew Hogan. 2021. The $162 Billion Design Industry Won't Stop Growing. https://www.forbes.com/sites/forrester/2021/03/25/the-162-billion-design-industry-wont-stop-growing/

[39] Niklas Holzner, Sebastian Maier, and Stefan Feuerriegel. 2025. Generative AI and Creativity: A Systematic Literature Review and Meta-Analysis. *arXiv preprint arXiv:2505.17241* (2025).

[40] James A Horne. 1988. Sleep loss and "divergent" thinking ability. *Sleep* 11, 6 (1988), 528–536.

[41] Stacy Hsueh, Marianela Ciolfi Felice, Sarah Fdili Alaoui, and Wendy E. Mackay. 2024. What Counts as 'Creative' Work? Articulating Four Epistemic Positions in Creativity-Oriented HCI Research. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 497, 15 pages. doi:10.1145/3613904.3642854

[42] Younjung Hwang and Yi Wu. 2025. The influence of generative artificial intelligence on creative cognition of design students: a chain mediation model of self-efficacy and anxiety. *Frontiers in psychology* 15 (2025), 1455015.

[43] Kyle E Jennings. 2010. Developing creativity: Artificial barriers in artificial intelligence. *Minds and Machines* 20 (2010), 489–501.

[44] Youngseung Jeon, Seungwan Jin, Patrick C Shih, and Kyungsik Han. 2021. FashionQ: an ai-driven creativity support tool for facilitating ideation in fashion design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[45] Daniel Kahneman. 2011. *Thinking, fast and slow*. macmillan.

[46] 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 preprint arXiv:1906.10188* (2019).

[47] James C Kaufman and Robert J Sternberg. 2010. *The Cambridge handbook of creativity*. Cambridge University Press.

[48] Chihiro Kawashima, Chong Chen, Kosuke Hagiwara, Tomohiro Mizumoto, Mino Watarai, Takaya Koga, Fumihiro Higuchi, Yuko Fujii, Emi Okabe, and Shin Nakagawa. 2024. Evaluating the impact of a short bout of stair-climbing on creative thinking in a between-subjects pretest posttest comparison study. *Scientific Reports* 14, 1 (2024), 176.

[49] Tae Soo Kim, DaEun Choi, Yoonseo Choi, and Juho Kim. 2022. Stylette: Styling the web with natural language. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[50] Claus-Peter Klas. 2012. Expert evaluation methods. *Dobreva et al* (2012), 75–84.

[51] Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. 2023. Large-scale text-to-image generation models for visual artists' creative works. In *Proceedings of the 28th international conference on intelligent user interfaces*. 919–933.

[52] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI? Design ideation with cooperative contextual bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[53] Barry Kudrowitz and Caitlin Dippo. 2013. Getting to the novel ideas: Exploring the alternative uses test of divergent thinking. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 55928. American Society of Mechanical Engineers, V005T06A013.

[54] Chinmay Kulkarni, Stefania Druga, Minsuk Chang, Alex Fiannaca, Carrie Cai, and Michael Terry. 2023. A word is worth a thousand pictures: Prompts as AI design material. *arXiv preprint arXiv:2303.12647* (2023).

[55] Harsh Kumar, Jonathan Vincentius, Ewan Jordan, and Ashton Anderson. 2025. Human Creativity in the Age of LLMs: Randomized Experiments on Divergent and Convergent Thinking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 23, 18 pages. doi:10.1145/3706598.3714198

[56] Elisa Kupers, Marijn Van Dijk, and Andreas Lehmann-Wermser. 2018. Creativity in the here and now: A generic, micro-developmental measure of creativity. *Frontiers in psychology* 9 (2018), 2095.

[57] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User experience design professionals' perceptions of generative artificial intelligence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[58] Antonios Liapis, Georgios N Yannakakis, Constantine Alexopoulos, and Phil Lopes. 2016. Can computers foster human users' creativity. *Theory and praxis of mixed-initiative co-creativity. Digital Culture & Education* 8, 2 (2016), 136–153.

[59] Jannik Lidner. 2024. Key Adobe Photoshop Statistics Revealed: Dominance in Image Editing. https://wifitalents.com/statistic/adobe-photoshop/

[60] Darren M Lipnicki and Don G Byrne. 2005. Thinking on your back: solving anagrams faster when supine than when standing. *Cognitive Brain Research* 24, 3 (2005), 719–722.
[61] James Lloyd-Cox, Alan Pickering, and Joydeep Bhattacharya. 2022. Evaluating creativity: How idea context and rater personality affect considerations of novelty and usefulness. *Creativity Research Journal* 34, 4 (2022), 373–390.
[62] Yuwen Lu, Chengzhi Zhang, Iris Zhang, and Toby Jia-Jun Li. 2022. Bridging the Gap between UX Practitioners' work practices and AI-enabled design support tools. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.
[63] Luís Macedo and Amílcar Cardoso. 2002. Assessing creativity: the importance of unexpected novelty. *Structure* 1, C2 (2002), C3.
[64] Dag Øivind Madsen and Richard W Puyt. 2025. The 7Vs of AI Slop: A Typology of Generative Waste. *Available at SSRN 5558018* (2025).
[65] Mary Lou Maher. 2010. Evaluating creativity in humans, computers, and collectively intelligent systems. In *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*. 22–28.
[66] Karin Matsumoto, Chong Chen, Kosuke Hagiwara, Natsumi Shimizu, Masako Hirotsu, Yusuke Oda, Huijie Lei, Akiyo Takao, Yuko Fujii, Fumihiro Higuchi, et al. 2022. The effect of brief stair-climbing on divergent and convergent thinking. *Frontiers in behavioral neuroscience* 15 (2022), 834097.
[67] Midjourney.com. 2024. Midjourney. https://www.midjourney.com/home
[68] Caroline Mimbs Nyce. 2024. The Atlantic. https://www.theatlantic.com/technology/archive/2024/08/why-does-all-ai-art-look-same/679488/?utm_source=apple_news
[69] OpenAI. 2022. DALL·E 2. https://openai.com/index/dall-e-2/
[70] OpenAI. 2024. ChatGPT. https://chatgpt.com/
[71] Michelle Leanne Oppert, Valerie O'Keeffe, Markus Søbstad Bensnes, Alin L Grecu, and David H Cropley. 2023. The value of creativity: A scoping review. *Journal of creativity* 33, 2 (2023), 100059.
[72] Marily Oppezzo and Daniel L Schwartz. 2014. Give your ideas some legs: the positive effect of walking on creative thinking. *Journal of experimental psychology: learning, memory, and cognition* 40, 4 (2014), 1142.
[73] Andreas Ostermaier and Matthias Uhl. 2020. Performance evaluation and creativity: Balancing originality and usefulness. *Journal of Behavioral and Experimental Economics* 86 (2020), 101552.
[74] Vishakh Padmakumar and He He. 2023. Does Writing with Language Models Reduce Content Diversity? *arXiv preprint arXiv:2309.05196* (2023).
[75] Srishti Palani and Gonzalo Ramos. 2024. Evolving roles and workflows of creative practitioners in the age of generative AI. In *Proceedings of the 16th Conference on Creativity & Cognition*. 170–184.
[76] Jonathan A Plucker, Ronald A Beghetto, and Gayle T Dow. 2004. Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational psychologist* 39, 2 (2004), 83–96.
[77] Xuebing Qin and Zhibin Jiang. 2019. The impact of AI on the advertising process: The Chinese experience. *Journal of Advertising* 48, 4 (2019), 338–346.
[78] Qualtrics. 2024. Qualtrics XM: The Leading Experience Management Software. https://www.qualtrics.com/
[79] Katharina Reinecke, David R Flatla, and Christopher Brooks. 2016. Enabling designers to foresee which colors users cannot see. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2693–2704.
[80] Katharina Reinecke and Krzysztof Z Gajos. 2014. Quantifying visual preferences around the world. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 11–20.
[81] Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1364–1378.
[82] Katharina Reinecke, Tom Yeh, Luke Miratrix, Rahmatri Mardiko, Yuechen Zhao, Jenny Liu, and Krzysztof Z Gajos. 2013. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2049–2058.
[83] Christian Remy, Lindsay MacDonald Vermeulen, Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2020. Evaluating Creativity Support Tools in HCI Research. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 457–476. doi:10.1145/3357236.3395474
[84] Mitchel Resnick, Brad Myers, Kumiyo Nakakoji, Ben Shneiderman, Randy Pausch, Ted Selker, and Mike Eisenberg. 2005. Design principles for tools to support creative thinking. (2005).
[85] Mel Rhodes. 1961. An analysis of creativity. *The Phi delta kappan* 42, 7 (1961), 305–310.
[86] Mark A Runco. 1993. Divergent thinking, creativity, and giftedness. *Gifted Child Quarterly* 37, 1 (1993), 16–22.
[87] Mark A Runco. 2007. To understand is to create: An epistemological perspective on human nature and personal creativity. (2007).
[88] Mark A Runco and Garrett J Jaeger. 2012. The standard definition of creativity. *Creativity research journal* 24, 1 (2012), 92–96.
[89] Mark A Runco and Shawn M Okuda. 1988. Problem discovery, divergent thinking, and the creative process. *Journal of youth and adolescence* 17, 3 (1988), 211–220.
[90] Ben Shneiderman. 2002. Creativity support tools. *Commun. ACM* 45, 10 (oct 2002), 116–120. doi:10.1145/570907.570945
[91] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (2007), 20–32.
[92] Robert J Sternberg and James C Kaufman. 2018. *The nature of human creativity*. Cambridge University Press.
[93] Robert J Sternberg, Jacques Lautrey, and Todd I Lubart. 2003. *Where are we in the Field of Intelligence, How did we get here, and Where are we Going?* American Psychological Association.
[94] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The metacognitive demands and opportunities of generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–24.
[95] E Paul Torrance. 1966. Torrance tests of creative thinking. *Educational and psychological measurement* (1966).
[96] E Paul Torrance. 1972. Career patterns and peak creative achievements of creative high school students twelve years later. *Gifted Child Quarterly* 16, 2 (1972), 75–88.
[97] E Paul Torrance. 1981. Predicting the creativity of elementary school children (1958-80)—and the teacher who" made a difference". *Gifted Child Quarterly* 25, 2 (1981), 55–62.
[98] Severi Uusitalo, Antti Salovaara, Tero Jokela, and Marja Salmimaa. 2024. "Clay to Play With": Generative AI Tools in UX and Industrial Design Practice. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1566–1578. doi:10.1145/3643834.3661624
[99] Veera Vimpari, Annakaisa Kultima, Perttu Hämäläinen, and Christian Guckelsberger. 2023. "An Adapt-or-Die Type of Situation": Perception, Adoption, and Use of Text-to-Image-Generation AI by Game Industry Professionals. *Proc. ACM Hum.-Comput. Interact.* 7, CHI PLAY, Article 379 (Oct. 2023), 34 pages. doi:10.1145/3611025
[100] James Vincent. 2022. "An engine for the imagination": the rise of AI image generators. https://www.theverge.com/2022/8/2/23287173/ai-image-generation-art-midjourney-multiverse-interview-david-holz
[101] Samangi Wadinambiarachchi, Ryan M Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The effects of generative ai on design fixation and divergent thinking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.
[102] Chetan Walia. 2019. A dynamic definition of creativity. *Creativity Research Journal* 31, 3 (2019), 237–247.
[103] Peter West, Ximing Lu, Nouha Dziri, Faeze Brahman, Linjie Li, Jena D Hwang, Liwei Jiang, Jillian Fisher, Abhilasha Ravichander, Khyathi Chandu, et al. 2023. The Generative AI paradox:" What it can create, it may not understand". *arXiv preprint arXiv:2311.00059* (2023).
[104] Yihan Wu and Wilma Koutstaal. 2020. Charting the contributions of cognitive flexibility to creativity: Self-guided transitions as a process-based index of creativity-related adaptivity. *PloS one* 15, 6 (2020), e0234473.
[105] Bin Yang, Long Wei, and Zihan Pu. 2020. Measuring and improving user experience through artificial intelligence-aided design. *Frontiers in Psychology* 11 (2020), 595374.
[106] Georgios N Yannakakis, Antonios Liapis, and Constantine Alexopoulos. 2014. Mixed-initiative co-creativity. (2014).
[107] Teng Ye, Katharina Reinecke, and Lionel P Robert Jr. 2017. Personalized feedback versus money: The effect on reliability of subjective data in online experimental platforms. In *Companion of the 2017 ACM conference on computer supported cooperative work and social computing*. 343–346.
[108] Weitao Zhang, Zsuzsika Sjoerds, and Bernhard Hommel. 2020. Metacontrol of human creativity: The neurocognitive mechanisms of convergent and divergent thinking. *NeuroImage* 210 (2020), 116572.
[109] Chuyi Zhou, Xiyuan Zhang, and Chunyang Yu. 2023. How does AI promote design iteration? The optimal time to integrate AI into the design process. *Journal of Engineering Design* (2023), 1–28.
[110] Eric Zhou and Dokyun Lee. 2024. Generative artificial intelligence, human creativity, and art. *PNAS nexus* 3, 3 (2024), pgae052.
[111] Feng Zhu and Wenbo Zou. 2023. The Role of Generative AI in Human Creative Processes: Experimental Evidence. *Preprint, Available at SSRN 4676053* (2023).

Table 2: Participant design and AI background, including self-evaluated design experience, typical design outputs, prior AI tools experience for design purposes, duration of AI usage (marked as N/A for participants who have never used AI tools for design), and self-reported AI experience.

PID	Design Experience	Design Output Type	AI Tools Used for Design	Duration of AI Use
*P*1	Beginner	Posters,3-D sculptures.	Never used	N/A
*P*2	Beginner	Posters,websites,apps	Never used	N/A
*P*3	Beginner	Posters,layout and publication design	Midjourney,ChatGPT, Bing	3 - 6 month
*P*4	Intermediate	Posters,advertising,websites,branding design	Never used	N/A
*P*5	Expert	Posters,advertising,websites,layout and publication design,branding design	Midjourney	3 - 6 month
*P*6	Intermediate	Posters,advertising,websites,layout and publication design,branding design	ChatGPT	3 - 6 month
*P*7	Minimal	Posters	Never used	N/A
*P*8	Advanced	Posters,websites,layout and publication design	ChatGPT	6-12 month
*P*9	Beginner	Posters,advertising,websites,branding design	Midjourney,ChatGPT	2 -3 month
*P*10	Beginner	Websites	Never used	N/A
*P*11	Beginner	Posters,advertising,websites,branding design	Never used	N/A
*P*12	Intermediate	Posters,advertising,websites,layout and publication design,branding design	Midjourney,ChatGPT	3 - 6 month
*P*13	Advanced	Posters,websites,layout and publication design	ChatGPT	6-12 month
*P*14	Beginner	Posters,websites	Never used	N/A
*P*15	Advanced	Posters,advertising,websites,layout and publication design,branding design	Midjourney,ChatGPT	3 - 6 month
*P*16	Intermediate	Poster, architecture	ChatGPT	6-12 month
*P*17	Beginner	Posters,websites	Stable Diffusion,ChatGPT	6-12 month
*P*18	Beginner	Posters	Never used	N/A
*P*19	Beginner	Posters,websites	Never used	N/A
*P*20	Expert	Posters,advertising,websites,layout and publication design,branding design,App	Midjourney,ChatGPT	6-12 month
*P*21	Intermediate	Posters,websites,layout and publication design	Midjourney,ChatGPT	3 - 6 month
*P*22	Beginner	Posters,websites	Never used	N/A
*P*23	Beginner	Posters	Never used	N/A
*P*24	Intermediate	Posters,websites,layout and publication design	Midjourney,ChatGPT	3 - 6 month
*P*25	Advanced	Posters,websites,layout and publication design	Midjourney,ChatGPT	3 - 6 month
*P*26	Intermediate	Posters,websites,layout and publication design	ChatGPT	Just started
*P*27	Beginner	Posters,websites	Never used	N/A
*P*28	Beginner	Posters	Never used	N/A
*P*29	Expert	Posters,websites,layout and publication design	Midjourney,ChatGPT	3 - 6 month
*P*30	Beginner	Posters,websites	Never used	N/A
*P*31	Intermediate	Posters,websites,layout and publication design,branding design	ChatGPT,Other,Microsoft Designer,Bing	6-12 month
*P*32	Intermediate	Posters,advertising,websites,layout and publication design,branding design	ChatGPT	< 1 month
*P*33	Advanced	Posters,advertising,websites,branding design	Midjourney,ChatGPT	6-12 month
*P*34	Minimal	Posters,advertising,websites,layout and publication design,branding design	ChatGPT	6-12 month
*P*35	Beginner	Posters,advertising,layout and publication design	Midjourney,Gemini,ChatGPT	> 1 year
*P*36	Advanced	Posters,websites	ChatGPT	2 -3 month

Table 3: Participant demographics, including age group, gender, major or field of study, ethnicity, and native English speaker status.

PID	Age	Gender	Major / Field	Ethnicity	Native Speaker
*P*1	18-20	Female	Psychology	White	Yes
*P*2	21-29	Female	School of Design	Asian	No
*P*3	21-29	Female	Psychology	Multiple races	Yes
*P*4	18-20	Female	Information Technology	Asian	Yes
*P*5	50-59	Female	School of Design	Hispanic	Yes
*P*6	21-29	Female	CS	White	Yes
*P*7	18-20	Female	Biology	Hispanic	Yes
*P*8	21-29	Female	UX Design	Asian	Yes
*P*9	21-29	Female	UX Design	Asian	Yes
*P*10	21-29	Non-binary	Information Technology	Hispanic	Yes
*P*11	21-29	Female	Information Technology	Asian	Yes
*P*12	21-29	Female	UX Design	Asian	Yes
*P*13	21-29	Male	UX Design	White	Yes
*P*14	21-29	Female	UX Design	Asian	No
*P*15	21-29	Female	UX Design	Asian	Yes
*P*16	21-29	Female	UX Design	Asian	Yes
*P*17	21-29	Female	Information Technology	Asian	Yes
*P*18	21-29	Female	Psychology	Asian	Yes
*P*19	18-20	Female	UX Design	Asian	Yes
*P*20	21-29	Female	School of Design	Hispanic	Yes
*P*21	21-29	Female	School of Design	Hispanic	Yes
*P*22	18-20	Female	UX Design	Asian	Yes
*P*23	18-20	Female	Psychology	White	Yes
*P*24	30-39	Female	UX Design	White	Yes
*P*25	21-29	Female	UX Design	White	Yes
*P*26	21-29	Female	UX Design	Asian	Yes
*P*27	18-20	Female	Information Technology	Asian	Yes
*P*28	21-29	Female	Information Technology	Asian	Yes
*P*29	21-29	Female	CS	Asian	No
*P*30	18-20	Female	CS	White	Yes
*P*31	21-29	Non-binary	Art history	Asian	No
*P*32	21-29	Female	Information Technology	Asian	No
*P*33	21-29	Female	Information Technology	Asian	No
*P*34	21-29	Female	Applied Math	Asian	No
*P*35	18-20	Male	UX Design	Asian	Yes
*P*36	21-29	Female	UX Design	Asian	Yes

Table 4: Profiles of evaluators/clients who evaluated the design outcomes.

ID	Role	Background	Gender	Years involved with the lab (designs' target)
E1	Associate Professor	UX Design, Health Tech	Female	> 5
E2	Master Student	CS, UX Design	Male	1–2
E3	PhD Student	UX Design, Music	Female	3–5
E4	PhD Candidate	UX Design	Female	3–5
E5	PhD Student	UX Design	Male	3–5

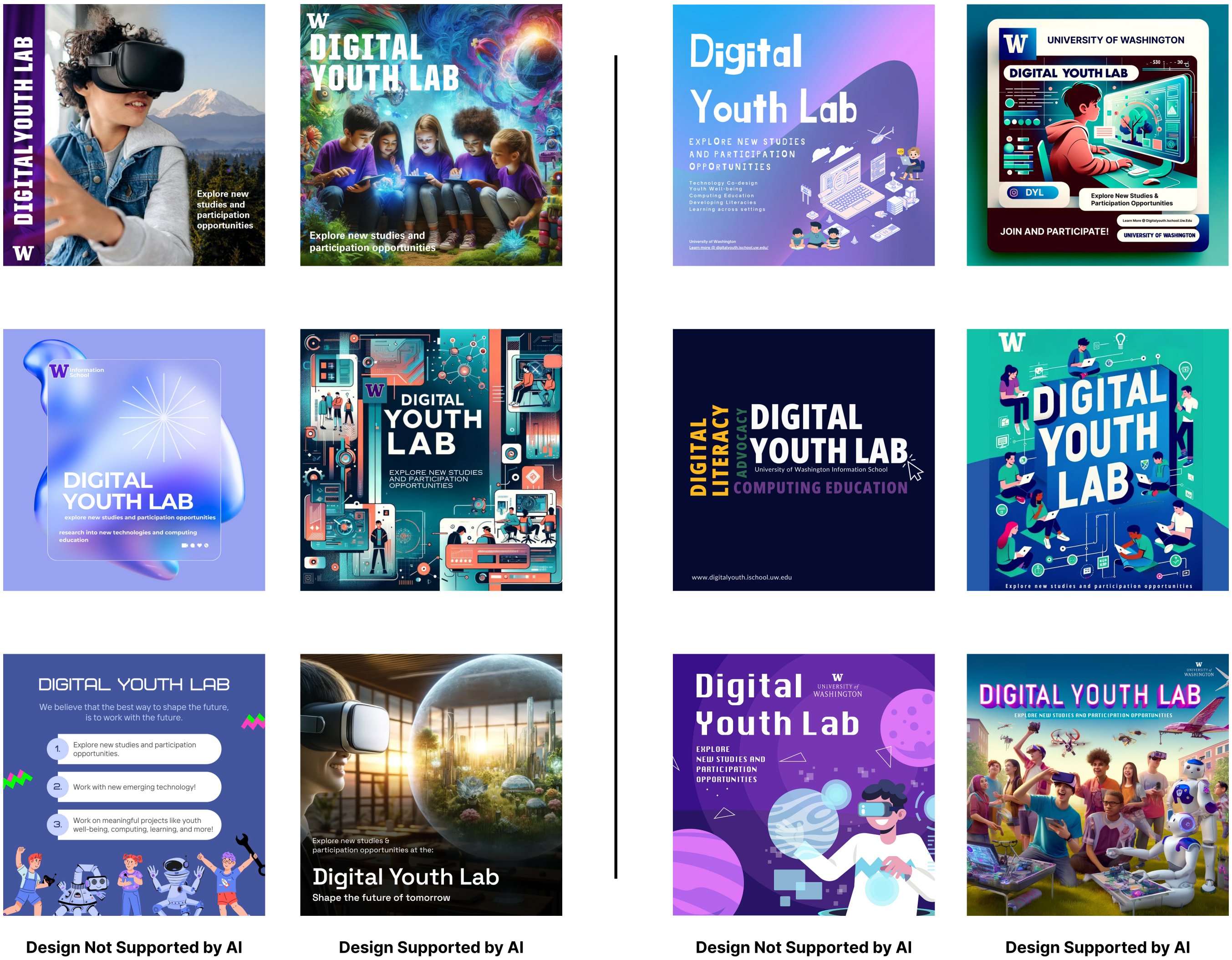

Figure 7: Additional examples of social media advertisement designs created by six participants. Each pair shows a design produced without AI support (left) alongside a design produced with AI support (right) by the same participant.