DesignPrompt: Using Multimodal Interaction for Design Exploration with Generative AI

Xiaohan Peng, Janin Koch, Wendy E Mackay

DesignPrompt: Using Multimodal Interaction for Design Exploration with Generative AI

Xiaohan Peng
xiaohan.peng@universite-paris-saclay.fr
Université Paris-Saclay, CNRS, Inria
Paris, France

Janin Koch
janin.koch@inria.fr
Université Paris-Saclay, CNRS, Inria
Paris, France

Wendy E. Mackay
mackay@lri.fr
Université Paris-Saclay, CNRS, Inria
Paris, France

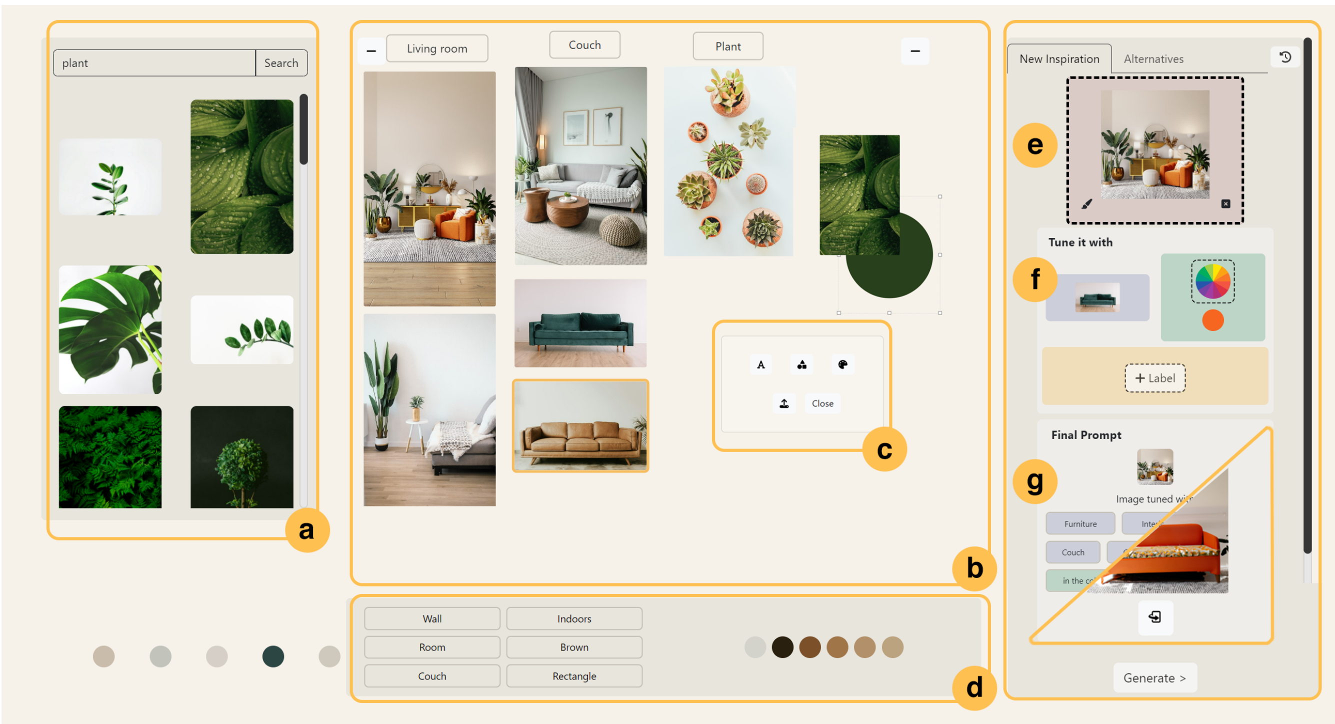


Figure 1: *DesignPrompt*: A digital moodboard tool that lets designers search images online (a) or generate AI images to create a moodboard (b) using common tools (c) as well as additional semantic meta-data of the moodboard images (d). Designers can compose multimodal GenAI prompts with images (e), colors, semantics and text (f) and finely tune their intentions (g).

ABSTRACT

Visually oriented designers often struggle to create effective generative AI (GenAI) prompts. A preliminary study identified specific issues in composing and fine-tuning prompts, as well as needs in accurately translating intentions into rich input. We developed *DesignPrompt*, a moodboard tool that lets designers combine multiple modalities — images, color, text — into a single GenAI prompt and tweak the results. We ran a comparative structured observation study with 12 professional designers to better understand their intent expression, expectation alignment and transparency perception using *DesignPrompt* and text input GenAI. We found that multimodal prompt input encouraged designers to explore and express themselves more effectively. Designer's interaction preferences change according to their overall sense of control over the GenAI and whether they are seeking inspiration or a specific image. Designers developed innovative uses of *DesignPrompt*, including developing elaborate multimodal prompts and creating a multimodal prompt pattern to maximize novelty while ensuring consistency.

DIS '24, July 1–5, 2024, IT University of Copenhagen, Denmark

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; • **Applied computing** → *Arts and humanities.*

KEYWORDS

Creativity Support Tool; Generative AI; Design Practice; Human-AI Interaction; Human-AI Ideation; Moodboard

ACM Reference Format:
Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2024. *DesignPrompt*: Using Multimodal Interaction for Design Exploration with Generative AI. In *Designing Interactive Systems Conference (DIS '24), July 1–5, 2024, IT University of Copenhagen, Denmark*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3643834.3661588

1 INTRODUCTION

Generative AI (GenAI) lets designers create high-quality synthetic images using descriptive text prompts [24, 55]. Text-to-image AI applications like DALL·E 3 [51], stable diffusion [48] and midjourney [47] enable designers to generate novel images or iterate their visual design. GenAI also produces a wide range of creative output beyond images. In music composing, GenAI can produce consistent and high-quality music with text input [2]. In 3D graphics creation, GenAI can produce 3D assets using solely text or 2D images without requiring 3D models [53]. In dance practice, GenAI can process motion capture data and generate unfamiliar and non-realism movement patterns that benefit dance practitioners [3].

Despite GenAI's ease of access with text and creative potential, it still poses challenges for non-AI professionals [38, 70]. More specifically, Subramonyam et al. [60] identified three key challenges when engaging with Large Language Models (LLMs): the intentionality gap, where the user communicates goals without additional cognitive processes; the capability gap, where users have limited understanding of AI's abilities; and the language gap, where the user struggles to articulate themselves clearly and effectively to the AI. Apart from the 'language gap' that raises challenges for non-AI experts in general, the domain-specific challenges for designers are especially reflected in intentionality and capability: The creative process is simply substituted by the AI system, which offers no rationales for its generated outputs [43]. One of the few workarounds allowing designers to regain control is generating multiple outputs for comparison [43], which however, can be inefficient and disruptive to creative practice.

Another widely used technical solution to these challenges is "prompt engineering", which is a means of structuring text prompts to efficiently interact with AI models and guide the generation process toward desired results [57, 67]. Recent research on prompt engineering focuses on effective strategies for prompt writing [11, 39, 40, 52], interactive prompt feedback [16, 56] and customized suggested prompt refinement [5].

While these studies have offered insights into refining text prompts to generate more predictable results, for visually oriented designers, optimizing textual prompts in a rule-based way still seems counter-intuitive, especially in divergent processes such as visual ideation [36]. Apart from prompt engineering, there seems to be a lack of research on designerly approaches for interacting with GenAI. This raises a key challenge: How can we design the interaction with AI in a way that aligns more with design thinking, enabling designers to better leverage the creative capacities of GenAI?

In this research, we use moodboard as a design medium to study how GenAI can be incorporated into the design process in a meaningful way. Moodboard is *"a visual collage composed of images, text, and objects, created to aid in exploring and defining ideas related to the given design task"* [33, p. 37]. It is commonly used in design and fashion as it helps designers to visually express ideas that are difficult to elaborate verbally [10]. Besides, moodboards have shown the potential to serve for much more roles in design practice, such as framing, aligning, paradoxing, abstracting, and directing [44]. Adding to existing image composition on moodboards, GenAI images could come as a new source of inspiration. As reflected in the practice of Design Workbook — image collages possess suggestive power lying in combination and contrast; "hybrid images" (existing images combined with diagrams or renderings) mix real and imaginary [22]. AI-generated images have encompassed features from both "hybrid images" and image collages. Since moodboarding is also a rather "hybrid" process encompassing divergent and convergent workflows, we recognize the potential of using AI-generated images for professional moodboard practice.

In this paper, we presented *DesignPrompt*, a digital moodboard tool allowing multimodal input including images, colors and semantics to help designers explore and express their intentions better with existing GenAI models. Specifically, designers can (1) expressively build their prompt to AI with images, colors and semantics; (2) explore connections and themes between visual inspirations by focusing on certain inputs; (3) review and revise prompts interactively before image generation; and (4) appropriate features based on their convergent-divergent needs.

The key contributions of this paper include: (1) a preliminary study that investigates how general audiences use GenAI applications for moodboarding resulting in four design implications for designing GenAI for that context; (2) *DesignPrompt*, a GenAI-powered moodboard tool with designer-centered multimodal input design; and (3) our insights based on a study with 12 professional designers. We conclude with a discussion and directions for future research in a broader context beyond moodboard.

2 RELATED WORK

2.1 Creativity Support Tools for Design Ideation

Creativity support tools (CSTs) are digital aids "encompassing one or more creativity-focused features" [20]. They support different phases in creative processes, with ideation or idea generation being one of the most commonly facilitated stages [14]. Ideation involves generating and developing novel ideas, which is closely related to divergent thinking in creativity research [30]. CSTs have been developed to support ideation across various means of interactions. IdeaWall and Idea Expander support multi-user brainstorming and collaborative creative meeting with conversation to visuals CSTs [58, 66]; GroupMind supports multi-user ideation with visual mind mapping CST [59]; MayAI supports human-computer co-design ideation using cooperative contextual bandits [34].

In design ideation in particular, designers navigate the iterative processes between analysis and synthesis of ideas or concepts in order to formulate a potential future [61]. Ideation in the context of design can be ambiguous and abstract, however, such ambiguity and abstraction are aligned with "designerly" thinking [18, 35]. Visual material is particularly suitable for such processes due to its high expressiveness and informational richness [61].

Tied closely with design practice, most CSTs are developed for design usages [20]. CSTs have been developed to support design processes in different design disciplines such as user interface (UI) design [49], fashion design [28] and graphical design [27, 29] and others. Since AI has exhibited creative potential particularly in finding diverse visual inspiration, there is a growing trend of integrating AI into design ideation CSTs. For instance, *SemanticCollage* [36] has a semantic search feature that uses AI-recognized semantic labels to support both visual and semantic-based ideation. Their work shows that semantic labels can support more holistic design exploration and reflection within the ideation process. *ImageSense* further combines semantic search and AI-suggested images in a collaborative moodboard [35]. In their work, designers build digital moodboards in a human-human-AI collaborative setting, while the AI analyses the moodboard content to find new (existing) image suggestions. While this was shown to support idea exploration with AI as a collaborator, it does not support design practitioners to express themselves visually. We see an opportunity to use a more generative approach to help designers not only search but also create visual material that aligns with their intention.

Recent advancements in GenAI allow novel synthetic image generation, which can push CSTs to a new boundary. We foresee that more CSTs will emerge as "generator" [26] and "design material" [19] in the design process. GenAI usually requires specification such as descriptive text prompting in order to produce meaningful and inspiring results. However, such requirements limit designers' ability to think visually and abstractly in design ideation. This suggests rather than simply "inserting" existing GenAI interfaces in CSTs, the integration of GenAI into CSTs should support visual and abstract thinking to make the combination inspiringly meaningful to design ideation.

2.2 Generative AI for Design Practice

One of the most common creative usages of GenAI is AI image synthesis, where Generative Adversarial Networks (GANs) and diffusion models both play important roles currently. GAN is known for its strengths in generating high-fidelity images [7]. However, with the development of diffusion models and LLMs, current image generative models such as DALL·E [51], stable diffusion [48] and midjourney [47] have outperformed GAN especially in image diversity and allow a broader audience to generate high-quality stylized images based on textual descriptions [17].

Currently, GenAI has already been applied in a wide range of design practices such as UI design of different fidelities [49], design concept exploration [65], fashion style clustering, forecasting and merging [28], speeding up design to avoid unnecessary tasks [62], ad-hoc exploration of visual ideas [64], visual communication and fast prototyping [12, 32]. However, we also notice in many creative AI applications, particularly those focused on result-oriented designs, AI usually replaces rather than enhances the creative process [64], thereby restricting designers' influence on the overall creative process. While there is a large body of literature describing the potential of GenAI for design professionals based on algorithmic capabilities, only a few works have investigated how GenAI can be adapted and meaningfully controlled to suit more diverse professional design processes [64].

In computer vision community, there is an increasing focus on how to make GenAI more controllable. While not directly applicable to design practice, works such as DreamBooth [54], ControlNet [71] and InstructPix2Pix [8] have allowed users to further edit and specify more image details using text prompts. Design workflows such as using ComfyUI [15] combined with ControlNet [37] have provided insights on how stable diffusion-based applications can be built more controllable and usable. However, workflows as such can still be seen as segmented end-to-end processes.

Focusing on GenAI's capabilities and guiding design with generated outcomes barely facilitates designing with GenAI meaningfully [56, 62]. Making sense of GenAI in professional design practice necessitates engaging designers in the decision-making process and using the technology in more realistic and professional settings.

2.3 Challenges of AI Prompting in Creative Processes

GenAI has posed challenges to non-AI experts to generate meaningful results due to difficulties in writing effective prompt input, uncertainties of AI's capabilities, insufficient control and the rather "black-box" interface [9, 38, 60, 68–70]. To make GenAI more relevant in professional creative practice, several existing works have focused on optimizing or guiding prompt engineering in image generation for creative purposes [5, 11, 40]. Chiou et al. have conducted a study on the co-ideation between participants and AI image generators, based on which generalized a few key strategies for generating effective prompts in design such as preparing data beforehand, breaking design into steps, crafting specific styles and reviewing to improve. However, apart from seeing possible challenges arising from altering the design process, we can also find the notion of "effective prompt" here is far from the natural language used by general audience. One might argue additional learnings are required to master prompting skills; nevertheless, interacting with such systems has become closer to instructing the system [62] iteratively with trials and errors [56], which can be challenging in a creative context.

Apart from prompt engineering, novel ways of interacting with AI have emerged as new opportunities in helping creative professionals utilize AI creatively and efficiently beyond text prompts. For instance, inpainting and outpainting allow users to input an image and modify or expand it with text [51]; multimodal guided artwork diffusion model allows synthesizing digital art differently by using text and image input in different weights [25]; 3DALL-E allows text and image input to generate 3D models [41]; PromptPaint allows interaction with a combination of text prompts and paint mediums such as oil painting and watercolor [13].

Reflecting on current design industry standards such as Adobe Firely [1], we can see despite features such as "Text to Image", "Generative Fill" and "Text Effects" seem to be able to produce aesthetic results that meet professional standards, interaction-wise these features all require detailed text prompting, which still seems counter-intuitive and contradictory to fit in a designer tool mainly composed of graphical functionalities. Motivated by prior research on creative multimodal AI and steering the perspective to a more designer-centered interactive system, we argue that using multimodal interaction may be a more designerly way of interaction that encourages designers' visual inspiration through the use of GenAI.

GenAI is still new in the design context. Based on our investigations there seems to be a lack of previous studies on how designers can use such technology for visual design tasks, especially ideation. This includes questions such as which part of the design process would benefit and be elevated using GenAI. Therefore we conducted a preliminary study to support the design and development process of *DesignPrompt*.

3 PRELIMINARY STUDY

We conducted a preliminary study to better understand how visual thinkers could incorporate GenAI into their moodboard process. We focused on users' interaction and workflow with existing moodboard tools, and how they phrase GenAI text prompts for visual ideation and respond to the generated results.

Participants We recruited eight participants (5 women, 3 men; age 23-30). Four had design backgrounds (P1, P2, P6, P8) and four had engineering backgrounds. P1, P3, P5, P8 are currently HCI researchers. Participants self-reported different levels of familiarity with GenAI: novice (P2-5, P7), intermediate (P1, P8) and expert (P6). We obtained informed consent from all participants, according to our IRB and followed European privacy laws (GDPR) with respect to all data collected from the study. All participants agreed to screen-record the study and anonymized publication of their results.

Setup The study was conducted in person with a researcher in attendance. The researcher provides access to a Microsoft Windows laptop with valid credentials to all the software required by the study. The participant sits at a desk with the laptop, a large external monitor, a mouse and a keyboard. The researcher sits aside. The researcher launches participants' preferred moodboard tools e.g. Pinterest or Behance. Similarly, the researcher also launches participants' preferred text-to-image GenAI applications, e.g. DALL·E, stable diffusion or Midjourney.

Procedure The whole study lasts 30-40 minutes in total. After welcoming the participant, the researcher describes the study and obtains informed consent. Next, the researcher describes the study design, which consists of two 12-minute tasks: a moodboard baseline task and a GenAI task. Participants are asked to use a talk-aloud protocol to describe their strategies for finding images or creating prompts. The researcher offers a short tutorial for participants who lack experience either with moodboarding or using GenAI.

Moodboard task (baseline condition): Participants are asked to create a moodboard that expresses a visual concept for a café interior in the style of their choice. They may use any image search engine or moodboard tools they prefer.

Generative AI task (experimental condition): Participants are asked to generate images of Café exterior with their preferred GenAI tool.

To ensure consistency and reduce technical problems, the participant suggests a potential GenAI prompt to the researcher, who then enters it directly into the chosen tool. Once an image is generated, the participant either adopts that image for their moodboard, or asks the researcher to adjust the prompt. Participants may also choose to enter prompts themselves.

Data Collection We captured screen recordings during each task. We also collected hand-written notes made by the researcher while conducting the study. These notes contained participants' verbatim statements during the whole study.

Data Analysis We conducted an inductive thematic analysis [6] for the collected data. After reading through the collected data, we generated 10 initial codes. They are: *query term formulation, image abstraction, moodboarding workflow, combining elements, creating effective AI prompts, understanding AI results, controlling AI results, inspired by surprise, input richness,* and *input preferences.*

After examining these codes, we identified 4 emerged themes, which are related to how a system should support: (1) different levels of abstraction (2) translating intentions to rich inputs (3) identifying the impacts of prompts (4) controlling outputs engagingly.

For either step, one author was mainly responsible for the coding while co-authors sampled quotes, coded and discussed them along the process. Representative quotes and counts for impact scale are displayed to highlight shared insights between the participants.

3.1 Results

3.1.1 Formulating search terms and prompts: Most participants (6/8) used a top-down search approach during the first task. For example, P2 started her search with "LGBT cafe" which led her to search for "LGBT bar" and subsequently, she delved into detailed aspects such as "lighting," "food," and "decoration". However, in later stages, most participants struggled with coming up with search terms. P6 mentioned when she lacked inspiration for search terms, she input partial queries into the search bar and relied on autofill suggestions to explore relevant terms. In the second task, most participants (7/8) felt that the AI did not understand them. All the participants were not satisfied with the first image generated by the AI. P1 said, "I told AI 'no filling' in this part but it didn't understand."

3.1.2 Switching between focused search and design inspiration: Half of the participants (4/8) began with a preconceived style or concept that guided their moodboard search process. P7 mentioned having an initial idea of creating a modern cozy cafe before beginning the process. Others were inspired by the images they found. P1 found common themes, while P2 and P3 realized what they liked and would avoid in their design while exploring. Two participants found AI-generated images unexpected and surprising, but also appealing. When P1 was exploring the concept of an "oval-shaped wall", AI generated an image featuring an oval-shaped plant wall, which P1 did not anticipate but genuinely appreciated.

3.1.3 Decomposing images into their components: When participants were organizing the moodboard, 3 participants mentioned that they would love to somehow mark and merge the features they liked in pictures. P6 said he sometimes only liked a part of the picture, also he would like to "mix and match" to try how feasible a feature/element would be in the real design. During the Generative AI task, 2 participants also asked if they could have multiple images (which is not possible), or even the whole moodboard as input. P3 found the result with textual prompts quite limited. She said, "It would be super helpful if I could use all the images on the moodboard as input. The data from all these images will be so much richer than one single sentence."

3.1.4 Fine-tuning prompts to control the outcome: More than half of the participants commented with comparative phrases such as "too many plants" or "less colorful" regarding the generated images. Most prompt tuning based on this feedback failed as it is difficult to turn these comparative phrases into effective AI prompts. P4 tried to add "less green plants" directly in the prompt, however, still generated an image with a similar amount of green plants as before. Even straightforward descriptions as prompts proved ineffective such as altering furniture color using "black furniture" (P7), yet the outcome still contained colorful furniture. P1 said, "It is quite frustrating not being able to control what feature AI extracts."

3.1.5 *Shifting input preferences for search and image generation:*

From observing participants engaging in our two tasks, we found all users wrote different lengths of text as input when searching and generating images. When users searched images, they gave shorter search terms, usually within 3 words. When interacting with the GenAI instead they tended to use a more detailed description.

3.2 Design Implications

Based on the results of the previously described preliminary study, we identified four design implications (DIs), which will be discussed in the following sections.

DI 1: Systems should support search or prompt input using different levels of image abstraction and semantics. We found users have different input preferences such as using semantics or keywords. We also observed that users sometimes encounter difficulties in formulating effective search queries and prompt terms. These observations highlight the need to capture user's intents accurately. Therefore we draw an implication that the system should accommodate various levels of semantics and image abstraction to facilitate more precise expression of users' creative intentions.

DI 2: Systems should help users to translate their abstract intentions to richer prompts. We observed that users engaging with GenAI tend to provide rather expressive text and lack details as descriptive prompts. Some also want to express themselves using multiple and multimodal inputs such as several images. This highlights the need for a "translation" process that allows users to input abstract or even multimodal intentions and generate visually meaningful output.

DI 3: Systems should help users to identify the impact of prompts. Within our findings in the GenAI task, users expressed a desire for assistance in understanding and controlling the impact of prompts on the output generated by the system. These findings emphasize the importance of AI transparency and inform design of a need for systems to provide users with tools or mechanisms that help them identify and comprehend the influence of prompts on the generated output, empowering them to achieve the desired creative outcomes.

DI 4: Systems should allow users to control and manipulate images engagingly. During our observations, participants expressed a desire for the ability to manipulate and combine parts of images beyond relying solely on text prompts. We identify a lack of effective fine-tuning capabilities for specific objects in images, as well as a deficiency in the means to meaningfully merge images. To inform design from an interactive standpoint, we believe that the system's interaction should allow users to actively engage and manipulate in the image editing process, rather than merely replicating an image editing tool with GenAI.

4 DESIGN GOAL AND RESEARCH QUESTIONS

Based on these insights we developed *DesignPrompt*, a digital moodboard system powered by GenAI. We evaluated *DesignPrompt* to gain insights into the following research questions:

- RQ1: Does using multimodal input to create GenAI prompts allow designers to explore and express their intents better?
- RQ2: Does revealing system interpretation of user prompts help users to produce results that are more aligned with their expectations?
- RQ3: Does interactive and controllable GenAI input let users perceive the system as more transparent and useful for design practice?

We aimed in particular to facilitate prompting using different levels of image abstraction and semantics (DI 1) as well as to translate their abstract intentions to richer prompts (DI 2).

5 *DESIGNPROMPT*: SCENARIO

The following scenario illustrates how Amy, an interior designer, would use *DesignPrompt* to make a moodboard for a living room design. Amy launches *DesignPrompt*, and an interface containing a search engine, moodboard canvas, image meta-data display and AI image generation tool shows up as shown in Fig.2. She starts by searching keywords such as "living room", "cozy sofa" and "plant"

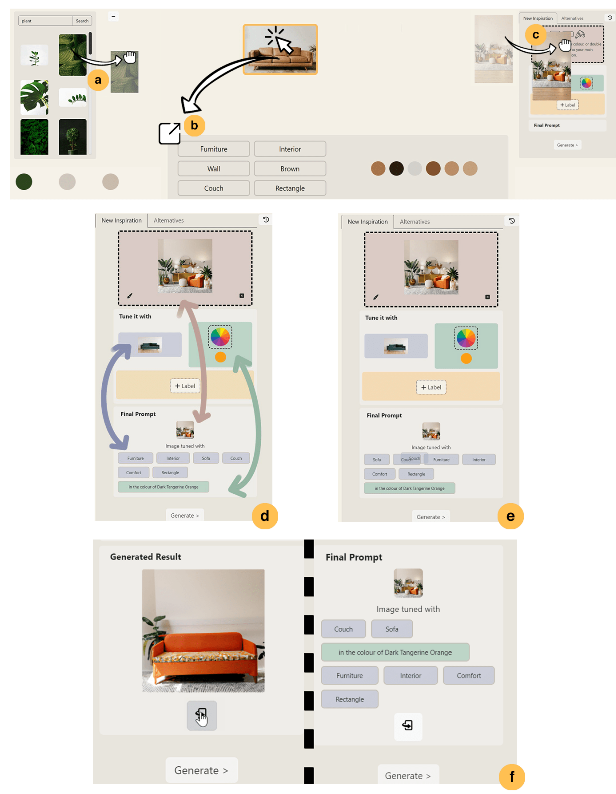


Figure 2: Interactions with *DesignPrompt*. Top: Drag image from search to canvas (a), click to open meta-data display (b), and drag elements from canvas to AI toolbar (c). Middle: AI toolbar maps multimodal input to final prompt (d) in real-time. The interactive prompt editing allows editing and reordering of prompt segments (e). Bottom: Generated results and final prompt (f) can be flipped.

at the search engine panel on the left (1), and scrolls to view more results. She drags and drops the images onto the canvas (Fig.2.a) and then resizes and rotates some images on canvas by double-clicking them. As Amy drags more images onto the canvas, the color palette at the bottom left of the canvas starts to build up. She single clicks on the sofa image and additional semantic and color information are displayed on the bottom (Fig.2.b), which she explores for new search ideas.

Amy then decides to generate images to refine her ideas. She starts by dragging an image of a living room from the canvas to AI "inspiration" box (Fig.2.c). In order to see how an orange couch would fit this room design, she drags the sofa picture and an orange color she likes to the tuning track (Fig.2.d). Amy reviews her prompt at the "Final Prompt" area. The prompt is color-coded thus Amy can see that she has tuned the original image with semantic labels from the couch image and the color she added (Fig.2.d). She then adjusted the prompt order by dragging and editing (Fig.2.e). After clicking "generate" an image is generated and replaces the previous "Final Prompt". Amy can click on the flipping button to "flip" between the generated output and the textual prompt input like a card (Fig.2.f). The generated image has the overall looks and stylings of the original living room, but with an added orange couch. She finds the result inspiring so she drags this image onto her moodboard canvas. A bit later, Amy is interested in alternative decoration for

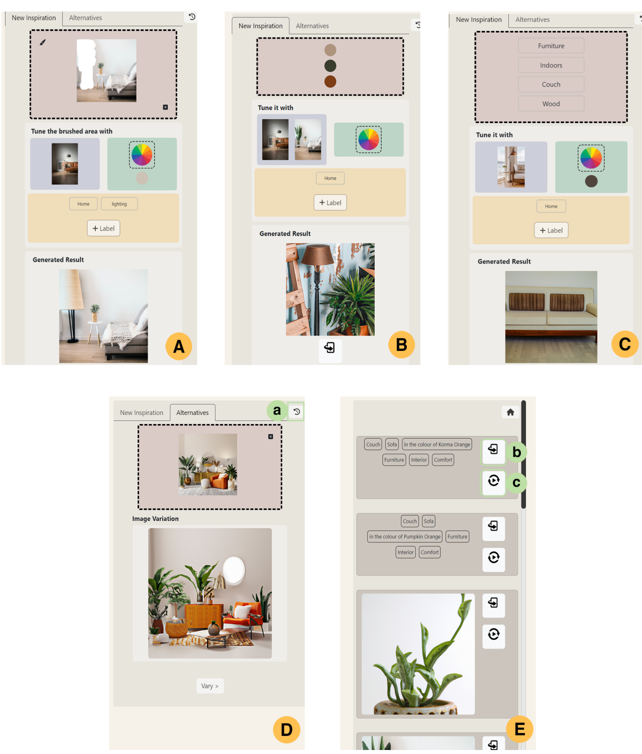


Figure 3: *DesignPrompt* allows multimodal input as inspiration, such as Image Inpaint (A), Colors (B) and Semantics (C). *DesignPrompt* has Image Variation Feature (D). *DesignPrompt* has Prompt History Revealing Feature (E).

the living room. She brushes the sofa in the inspiration box away as shown in Fig.3.A, and under instruction saying "Tune the brushed area with...", she drags a floor lamp picture, a beige color, as well as semantic labels saying "Home" and "lighting" to the tuning panel (yellow area) and generates a new image. Without having to type anything, Amy obtains an image of the living room with a beige floor lamp without any other changes, which is what she expects. She explores also other multimodal combinations focusing on color (Fig.3.B) and semantic concepts (Fig.3.C), or variation of an existing image (Fig.3.D). After a few generations, Amy reviews her idea trajectory by clicking on the "history" button (a) (Fig. 3.D) to view all the generated images (Fig. 3.E). She wants to see what prompt was used to create this image (Fig. 3.E) and wants to start her new search from there. So she reviews the generated image and prompt used by clicking button (b) (Fig. 3.E), then she clicks button (c) which reapplies the prompt to the AI image generation tool from which she continues her ideation process.

6 DESIGN AND IMPLEMENTATION

DesignPrompt is a digital moodboard system that uses GenAI for design exploration. The system is built as a web application that uses Vue3, a javascript front-end framework and Express.js, a node.js back-end framework. The application interface consists of these parts: search engine, moodboard canvas, canvas color palette, semantic and color meta-data display, an AI tool panel including GenAI multimodal input, alternative, interactive prompt editor, and GenAI history (Fig.4). All main interactions are through drag and drop, and all tools can be minimized to allow designers to concentrate on the moodboard itself. Inspired by SemanticCollage [36], *DesignPrompt* has a search engine, AI tool panel and an open canvas in the same interface, thus designers can freely and easily drag, drop and move images to switch between different features.

6.1 Search

We applied Unsplash, an open-source image API [63] for our search engine, which provides high-quality image searching functionalities and basic image information such as image tags (Fig.4.a). Images from search results can be dragged and dropped to canvas then reused in other tools.

6.2 Moodboard Construction Tools and Image Meta-Data Display

In order to allow a realistic study environment, we have added basic visual editing tools such as sizing, coloring as well as adding shapes and text to the moodboard canvas (Fig.4.b). Designers commonly use these tools to achieve a desired aesthetic and create professional and realistic moodboards [36]. Designers can double-click on the image itself to resize it and right-click on the canvas the trigger the "quick-add" popup on canvas. Designers can add customized text labels, colors, and shapes and upload their own images to canvas. A color palette shows all main colors of images on the canvas (Fig.4.c).

To assist designers in contextualizing and using images on the moodboard, *DesignPrompt* provides each image with meta-data consisting of a color palette and semantic labels (Fig.4.d). Single clicking on an image on canvas opens image meta-data display on the bottom of the canvas, which shows semantic labels obtained

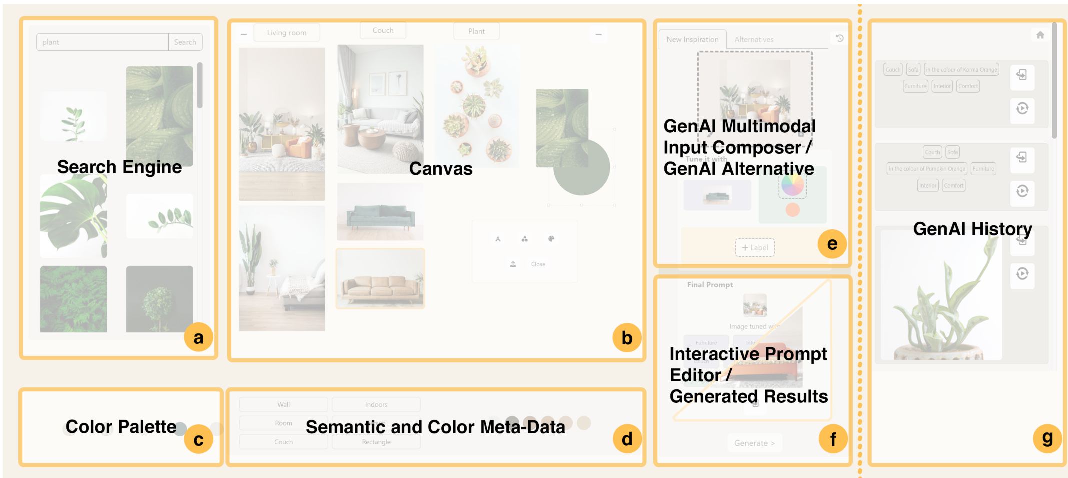


Figure 4: Overview of *DesignPrompt* structure including the search engine (a), moodboard canvas (b), canvas-based color palette (c), image-based semantic and color meta-data display (d), GenAI multimodal input composer and GenAI alternative (e), interactive prompt editor and generated results (f) and GenAI history (g).

from image recognition and original tags, as well as a color palette containing the six main colors recognized from the image. If two images are selected, the meta-data display will show the common semantic labels at the top, then the remaining semantic labels as well as the colors of each image.

Semantic Labels Similar to *SemanticCollage*[36], we used Google Vision API [23] to obtain the first six semantic labels for each moodboard image, which are attached in the order of recognition scores.

Color Palette We applied Colorthief [42], an open-source javascript feature to obtain the color palette of each image.

6.3 GenAI Tool Panel

Multimodal Input Composer for Image Generation *DesignPrompt*'s multimodal input composer accepts color, image, inpainted image, semantic labels, and text as inputs to help designers effectively convey their ideas. The composer requires input in one modal as inspiration, such as an image, some colors, text, or labels, as shown in the area (Fig. 4.e). Depending on how users would like to tune the inspiration, users can put multimodal combinations of inspiring colors, texts, labels, or images in the tuning area to guide and control image generation. Users can add these elements by dragging from canvas and dropping them in corresponding input areas. For tuning color specifically, apart from dragging an existing color recognized from images, users can also add a customized color by single clicking on a standard color wheel. We interpret those colors into a prompt which is composed of a color name and shade using the coding library "Name that Color" [46]. Users' image input in the tuning area is interpreted as prompts composed of semantic labels. For generating images we used the OpenAI API [50], which uses text prompts and image masks (optional) provided by users to generate, edit and get variations of images.

Interactive Prompt Editor In *DesignPrompt*, there is a direct mapping between multimodal input and machine-translated prompts, which are updated in real-time (Fig. 5.a). When tuning an inspiration with images, semantic labels of each image will be added to the "Final Prompt" (Fig. 4.f). The machine-translated color will be added using the prompt snipped: "in the color of xxx". Each tuning area is color-coded, as are the different elements of the prompt, allowing users easily recognize content origin. For instance, the semantic labels obtained from tuning images will be shown in purple, consistent with the color of the input location.

Regardless of different combinations of multimodal input, users can always view what exact input will be sent to AI under "Final Prompt" transparently as shown in Fig. 5.b, Fig. 5.c and Fig. 5.d. Additionally, *DesignPrompt*'s design uses design abstractions instead of AI terminology, allowing non-AI professionals to play with multimodal inspirations with clear instructions. The user interactions also remain consistent despite various modality choices. Users can specify a *dark green sofa* without specifying the type of *dark green*. Instead, they can simply drag *sofa* label or a picture of a sofa then drag a color to achieve their desired results.

Each prompt sequence in the "Final Prompt" section is interactive, allowing users to conveniently rearrange, edit, or extend them before initiating the generation process. After editing a prompt element, the mapping relationship between multimodal input and machine translation no longer exists, thus a new color code—blue will be assigned to the prompt sequence, representing "user translation" (Fig. 5.b). Clicking "Generate" button results in splicing all prompt sequences together separated by commas, and sending them as input to GenAI. When generation ends, the generated image replaces the interactive prompt. The prompt and image are added to the GenAI history. The 'flipping' button under the generated image allows users to flip back to view the prompts input used to generate the image as shown in Fig. 2.f.

GenAI History History feature allows designers to reflect and review both generation output and input (Fig.4.g). Designers can click 'flip' (button (b) in Fig.3.E) to flip between generated image and the exact prompts used for generation. The interaction is consistent with interactive prompt editor (Fig.2.f). Designers can reapply these input to GenAI by clicking "reapply" (button (c) in Fig.3.E).

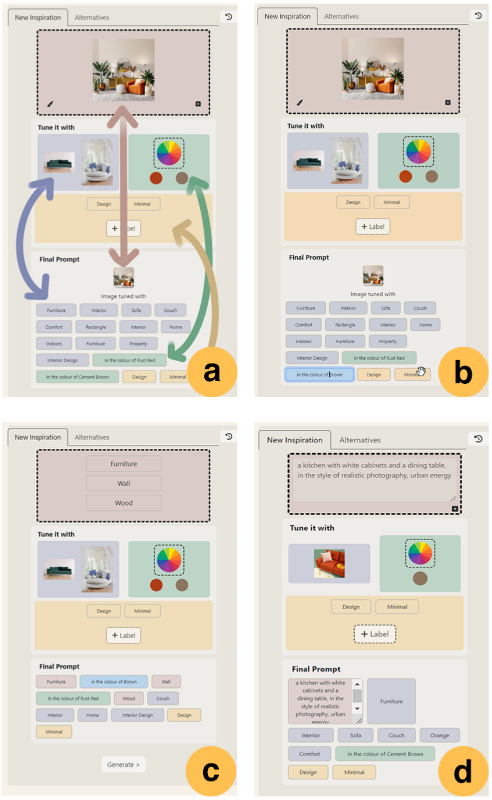
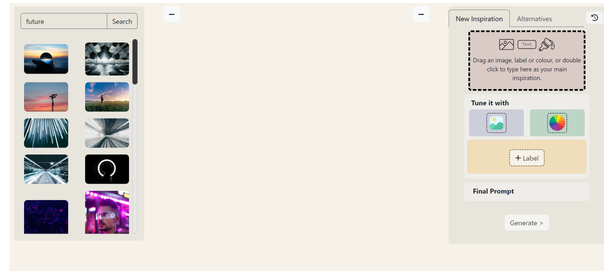


Figure 5: Interactive Prompt Editors allows mapping from multimodal input to text prompts (a), interactive prompt editing and reordering (b), interactive mapping with semantic labels as inspiration (c), interactive mapping with long text as inspiration (d).

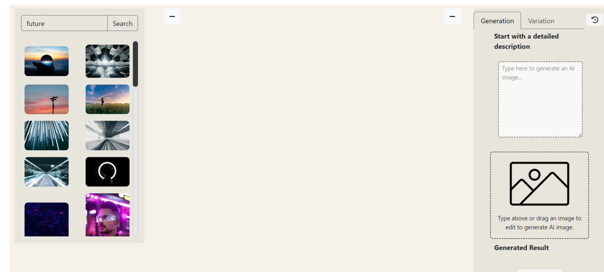
7 STUDY METHODOLOGY

We are interested in how standard text-based GenAI prompts compare to multimodal prompts, specifically: 1. What are the trade-offs between text-based and multimodal prompt strategies? 2. How do different prompt strategies align with participants' expectations? and, 3. How does the participant's interaction with different prompt strategies affect their perception of the efficacy of the system? We conducted a comparative structured observation study [45], similar to [4, 21, 31, 36]. This qualitative approach systematically controls the presentation of ecologically valid tasks that vary along one or more specified dimensions, in this case, *prompt strategy*. Participants perform multiple tasks with each design variant and then respond to questions that require them to reflect deeply upon and compare their strengths and weaknesses. An experimenter also directly observes how participants interact with each design variant. The goal is to understand the trade-offs across design variants from the user's perspective to evaluate and advance the design.

To provide a fair comparison, we created two variations of *DesignPrompt*, a moodboard creation system that differed only in the GenAI prompt interaction: The *Text condition* permits only text-based prompts and basic inpainting, similar to DALL·E's AI image



(a) ***Multi-modal condition***



(b) *Text condition*

Figure 6: The *Multi-modal condition* (a) and *Text condition* (b) each include a search engine (left panel), a central canvas, and a GenAI prompt space (right panel).

generation interface, and serves as the baseline. The *Multi-modal condition* permits multimodal prompts, including the above as well as images and color, and serves as the key design variant. At this writing, current moodbood tools do not integrate GenAI prompts and inpainting with an integrated search engine, and thus our baseline exceeds current capabilities. However, although this raises the bar for the baseline comparison, we also anticipate that these capabilities will arrive soon and so consider this a fairer comparison.

7.1 Participants

We recruited 12 participants (9 women, 3 men; aged 22-54) who have at least two years of design training or practice; and 2-20 years of moodboard design experience. Their experience with GenAI varied from novice to expert (see table 1). We followed our institution's IRB policy both in our explanations of the study to participants and study design. All participants signed informed consent forms and agreed to voice and screen recordings. We anonymized the results and followed European privacy law (GDPR).

7.2 Setup

A Microsoft Windows laptop running *DesignPrompt* sits on a table, with a large external monitor, a mouse and a keyboard. The experimenter launches the *Multi-modal condition* or the *Text condition* according to the experimental protocol.

7.3 Procedure

The study lasts approximately one hour. We use a [2x2] within-participant design with two factors. The primary factor related to PROMPT STRATEGY: *Text condition* (baseline, text-only prompts) and *Multi-modal condition* (variant with multimodal prompts). We

ID	Gender	Design Background	Profession	Moodboard Experience (yrs)	GenAI Experience (level)
1	M	Design Research	HCI Researcher	2	Intermediate
2	F	Graphic Design	Packaging Designer	8	Novice
3	F	Fashion Design	Product Designer	2	Novice
4	F	Architecture	HCI Researcher	2	Intermediate
5	F	Fashion Design	Fashion Designer	7	Novice
6	F	Fashion Design	Fashion Design Student	4	Novice
7	F	Media Informatics	HCI Student	2	Intermediate
8	M	Photography and Design	Publisher	20	Novice
9	F	Fashion Design	HCI Researcher	9	Intermediate
10	M	Media Design	HCI Researcher	10	Expert
11	F	Media Design	HCI Researcher	2	Intermediate
12	F	Industrial Design	UX Researcher	5	Intermediate

Table 1: Study participants' backgrounds and experience. (12 participants (9 women, 3 men; aged 22-54) with at least two years of design practice and 2-20 years of moodboarding experience, and various familiarities with GenAI from novice to expert).

created two equivalent DESIGN BRIEFS to generate the design tasks, with a Latin Square to counter-balance for order across conditions and participants. The study protocol consists of four steps:

(1) **Introduction and Tutorial:** The experimenter first describes the study and obtains informed consent. Participants then view a tutorial that describes how *DesignPrompt* works and have several minutes to familiarize themselves with different prompt strategies. The experimenter explains about GenAI to novice users and answers questions as required.
(2) **Design Tasks:** Participants perform two moodboard design tasks — *Text condition* or *Multi-modal condition* — with a think-aloud protocol to describe what they are doing and why. Each task begins with a short video tutorial with details about the current prompt strategy. Participants may ask the experimenter technical questions as needed. Participants first read the assigned DESIGN BRIEF and then perform a 12-minute moodboard task with the assigned prompt strategy: *Text condition* or *Multi-modal condition*. After answering a short questionnaire, they repeat the process with the second design task and the other prompt strategy.
(3) **Comparative Questionnaire:** The participant fills out a Likert-scale questionnaire to compare and contrast the two prompt strategies.
(4) **Comparative Interview:** The experimenter conducts a semi-structured interview (10-20 minutes) that asks participants to reflect upon the different prompt strategies. Some questions probe more deeply into aspects of the comparative questionnaire; others encourage participants to reflect on features common to both systems, such as their understanding of how prompts work. Participants are also asked to reflect upon how their role changes as they interact with GenAI and their view of its challenges and potentials.

The two design briefs are:

Design Brief A: Imagine you are organizing a meetup event with local designers, and you would like to have a space that not only inspires creativity but also fosters networking and collaboration. *Objective:* Create a moodboard to help you curate a dynamic and inspiring environment for that event.

Design Brief B: You are asked to redesign the shared kitchen space at work that hosts weekly "cake day" event. You would like to create a space that is both lively and comfortable for the team to enjoy cake and casual chats, also caters to colleagues from diverse backgrounds and professions. *Objective:* Create a moodboard to collect ideas for a vibrant yet cozy environment for the team's shared space.

7.4 Data Collection

We captured screen recordings during each task and recorded audio as the participants described their actions. We collected answers to the three Likert-scale questionnaires and two final moodboards created by each participant. We also collected handwritten notes made by the experimenter as they observed participants.

7.5 Data Analysis

We conducted a reflexive thematic analysis [6] using a mixed deductive (top-down) and inductive (bottom-up) approach for the recorded audio data, and also reported participants' answers to the comparative questionnaire.

First we analyzed and coded our collected data using a deductive approach guided by our research questions. The initial themes were system usefulness perception, feature usages and system appropriation, prompt building strategies, understanding of GenAI prompting, AI's impacts on users' roles, intention expression and exploration, idea convergence and professional usage potential.

Using an inductive approach, we reflected on patterns within and across conditions to gain a more nuanced understanding of the data. We also abstracted (sub-)themes based on the gained larger contextual knowledge. This process resulted in four main themes, each with several related sub-themes how: Multimodal input enables designers to explore and express their ideas (8.2), Designers interpreted "feeling in control" differently (8.3), Expectations mediated between designers' intentions and results (8.4), and how designers adapt to the system (8.5) when interacting with *DesignPrompt*. For either step, one author was mainly responsible for the coding while co-authors sampled quotes, coded and discussed them along the process. Representative quotes and counts are displayed to highlight shared insights between the participants.

8 RESULTS

This section provides an overview of questionnaire's quantitative results and qualitative findings from structured observations.

8.1 Questionnaire Results

The short questionnaires at the end of each design task show that both the *Text condition* and *Multi-modal condition* prompt strategies improve system expressivity, users' understanding of prompt and images' connection, and output understanding, and are less good at producing images that "align with expectations". This indicates that both the *Text condition* and *Multi-modal condition* prompt strategies are suitable for moodboard design.

The results of the comparative questionnaire (Fig.7) show that participants preferred the *Multi-modal condition* prompt strategy for exploration, expressivity, interaction worth the effort and user understanding on how prompts impact AI. However, they preferred the *Text condition* prompt strategy for ease of steering and generating images that makes sense. Participants preferred the *Multi-modal condition* strategy (41%) to *Text condition* (25%) to support ideation, but were mixed with respect to "feeling in control", with roughly equal numbers of participants preferring each.

8.2 Multimodal input enables designers to explore and express ideas

Most designers mentioned they enjoyed the multimodal input feature (8/12), especially in terms of expressing themselves more creatively (9/12) exploring different ideas (10/12) and providing them more options and opportunities (4/12).

Opening Up Creative Opportunities Participants enjoyed the creative options and opportunities *Multi-modal condition* provided (P2, P4, P7, P10). P1 also mentioned in the *Multi-modal condition* "was much better in terms of workflow" and "(make moodboarding) more enjoyable to do".

Expressive Power of Multi-Modalities "Decomposing" the input was particularly interesting as it allows"different things that you could feed and focus on", having "different steps (in constructing prompts)" (P9, P11), and expressing "more abstract and complicated concepts" (P2). Multiple modalities increased designers' expressive power in contrast to text prompts, which can make them rather "feel constrained" (P1). Designers found it difficult to formulate ideas into text (P4, P6), "Sometimes I don't even know what I want" (P6). Using images and colors as input "instead of writing them" (P9, P10) allowed designers to quickly express their ideas.

Exploring Connections between Inspirations Designers identified connections between inspirations and adapted their multiple modality input accordingly (P9, P12). P12 identified "decoration and purple" as key elements in her design, maintaining decoration-related semantics and color as persistent "links" in input composer. P9 liked how *Multi-modal condition* helped her find inspirations that are "blurry" and "in-between images" she connects to.

Encouraging Creative Thinking Participants reported the decomposed input motivated them to think diversely and creatively (P5, P7, P12). A fashion designer (P5) noted that the "*Multi-modal condition* aligns more with the way I think", allowing for diverse perspectives and exploration before narrowing down ideas. *Text condition* was perceived "more linear", while *Multi-modal condition* "elevated" to think more creatively (P12).

8.3 Designers interpreted "feeling in control" differently

While most designers preferred *Text condition* over *Multi-modal condition* in terms of ease of steering (8/12), others think interactive construction of the prompt in *Multi-modal condition* helps them to view the system as more transparent (8/12). One participant understood "control" as system interpretation, while 3/12 interpreted it as controlling input and again 4/12 interpreted as controlling over generation process.

Ease of Steering Some designers interpreted "control" as simplicity of understanding and use (P2, P3, P7). Most users became familiar with the *Text condition* rather quickly in our trial session regardless of their previous experience. Some found it easy to learn the *Multi-modal condition* thanks to our tutorial, while others took longer. P3 mentioned she was "more in control (with the *Text condition*)" because she thought it was "easier to use" than *Multi-modal condition*. P8 who rarely used AI prompting mentioned he felt more confident with the *Text condition* and the hierarchy in the *Multi-modal condition* made him "a bit lost".

System Interpretation Some designers struggled to understand the system's interpretation which affected their feeling of control (P3, P8). As an example, P3 put colors and images as input, but found the translation into AI-generated semantic labels confusing stating that "sometimes they got you prompts that you were not looking for", which made her feel out of control.

Controlling Input For some others, "control" means feeling control of their input (P5, P6, P9, P11). P5 mentioned with the *Multi-modal condition* she was able to "control precisely what I want" with images and colors.

Controlling over Generation Process "Control" can also mean control over the AI generation process for some designers (P1, P9, P11, P12). Most found *Multi-modal condition*'s interactive prompt editor made the GenAI system more transparent (8/12). The editing process facilitated designers to "control what information is fed into the AI model" (P11), like "reviewing the black box somehow" (P11). However, insufficient understanding of "how it (AI) combines images and takes what elements" (P12) can cause uncontrollable outcomes, which lead to increased "trials and errors". As workarounds, designer P1 used "at least a few generations" to deduct patterns, while P9 utilized the history feature to "reflect on the outcomes" and try to better understand the generation mechanism.

8.4 Expectations mediated between intentions and results

6/12 designers mentioned *Text condition* being able to produce "better" results than *Multi-modal condition*; 2/12 designers mentioned *Multi-modal condition* failed their expectations in illustrating precise ideas; 2/12 designers thought *Multi-modal condition* opens up

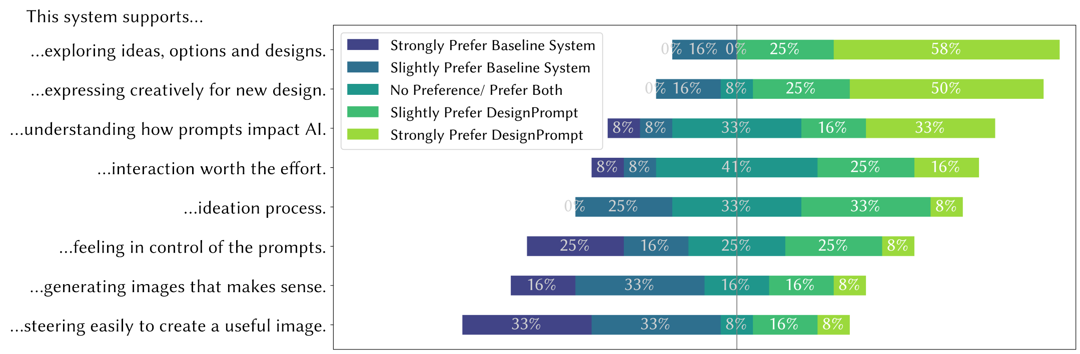


Figure 7: Comparative Questionnaire Results between our *Text condition* (blue) and *Multi-modal condition* (green).

their creativity; 2/12 designers enjoyed *Multi-modal condition* since it produced unique images creatively.

Interpretation of Surprises Several designers reported that *Text condition* produce "better" results than *Multi-modal condition* (P1, P2, P3, P8, P10, P11), which is supported by our questionnaires (see Fig.7). In the *Multi-modal condition*, designers got often "unexpected" (P2, P11) and "surprising" (P1) results. Some hypothesized that *Multi-modal condition* provided "too much information for the AI generation" (P2, P8) while *Text condition* provided "more detailed instruction" (P2). Unexpected results could also be caused by *Multi-modal condition* "encouraging" designers to put "an abstract idea" (P11) or "less precise prompts" (P1).

Precise Expectations Some designers had rather precise expectations in mind and were usually frustrated by the results. For instance, P7 "had a certain image I wanted to create, but it wasn't possible". P3 was further disappointed that the colors generated were not exactly what she put in the input.

Open Expectations Other designers were excited that AI-generated images they "cannot imagine" (P4, P9). P4 used image alternatives and found the new image "has the same room layout, but it adds lots of people and it shows a dynamic atmosphere. That's what I want, but I didn't imagine it before it happened." When asked about image predictability, P9 stated she was unsure if she was expecting or speculating out of curiosity. AI "quickly opened up perspectives" and intrigued her with the potential to mix them with her own ideas. She remained receptive to both the "AI's perspective" and its unpredictability. *Multi-modal condition* also benefited designers to generate unique images (P5, P12). P5 explained that at her design school "everyone ended up having a similar moodboard because everyone uses Pinterest and there are always a lot of pictures that everyone will use eventually", thus she appreciated the uniqueness enabling her to distinguish her ideas.

8.5 Designers adapt and also adapt to the system

3/12 designers mentioned a learning curve when using the *Multi-modal condition*; 4/12 designers mentioned the 12-minute moodboarding session was too short; 6/12 designers reported they better understood how prompts impacted AI after the study.

Adapting and Learning in Interaction Some designers reflected that the *Multi-modal condition* might have a "learning curve" and the time constraint of 12 minutes was too short for them to fully master it (P1, P3, P8, P11, P12). Designers adapted to the system and mentioned they would "find it more useful" after getting more used to it (P1, P12). At the same time, interacting with *Multi-modal condition* also impacted their design process. Both P4 and P10, who had prior experience with AI prompting, stated that it "affects the way I design" (P4) and "elevated" (P10) their creative process after learning more about the system. Designers who had little knowledge of prompting mentioned they "understand the key things (to AI input) now" (P6) and learned that GenAI "prioritizes the labels that I put at front" (P5) after using the interactive prompt construction feature in *Multi-modal condition*.

Appropriating Features Unexpectedly In the *Multi-modal condition*, most designers almost only used images as their base inspiration. However, designers used and appropriated features of *Multi-modal condition* in innovative and unexpected ways. Some designers input extremely long image prompts (P9, P12). P9 once tuned an image with 5 other images and she "couldn't really imagine an image based on all of that". P12 used a lengthy compound input with multiple modalities and quickly switched between base modalities. Additionally, P12 explored color as "a theme" rather than a descriptor in her design. She consistently kept the same color in all the following generations.

8.6 Other Findings

All the designers used image editing in their moodboarding process and 7 out of 12 participants complimented on using inpaintings in both systems that allow them to "change some details" (P2) and "customize" (P6) easily. 3/12 designers mentioned they enjoyed the feature of variation and they all preferred to have more variations generated at one time, and P8 suggested having it "running in the background" and "generating multiple images on the fly" based on the current moodboard.

9 DISCUSSION

9.1 Decomposing Prompts into Modalities

In our preliminary study users faced challenges when trying to articulate their intentions in text-based search terms or prompts. Some users also mention they would like to quickly try out a rich combination of different input modals like images and colors. In *DesignPrompt*, the decomposition of text input into colors, images and semantics enables more nuanced and articulated expression of intentions. This allows designers to go through different layers of their inspirations to express and connect them without explicit verbal descriptions. Designers also interpret the decomposition as a way of realizing different focuses and steps in their creative process.

In terms of exploration, apart from exploring new ideas from scratch with different combinations of modalities, separating them also enables designers to keep certain elements, such as color, consistent as input. This allows maintaining certain links between inspirations while exploring the middle ground between completely novel ideas and consistent themes. Since *DesignPrompt* also allows single-modal text input, it still offers flexibility for designers to express descriptively, as P4 said "if I first think of a text, then I use text. If I don't have a clear text and I just already have an image, then I use image."

Reflecting and extending on our RQ1, using multimodal input in *DesignPrompt* allows designers to explore different combinations of inspirations, which especially facilitates them in constructing the connections and themes between images. It also allows them to express their intents more flexibly compared to the *Text condition*.

9.2 Feeling in Control

Designers have rather mixed opinions about "feeling in control". Apart from the lack of control stemming from the unfamiliarity with how GenAI works in both *Multi-modal condition* and *Text condition*, around half of the designers "feel more control" in *Multi-modal condition* while the other half prefer *Text condition*. In terms of ease of steering particularly, they favored *Text condition* over *Multi-modal condition*.

Lack of control can stem from unfamiliarity with how GenAI works in general, such as how and what AI is combined and what causes it to fail. Designers often kept interacting with the system using similar patterns with trials and errors resulting in rather frustrating experiences. This is consistent with previous studies that demonstrated how designers often seek explanations and validate their hypotheses when interacting with such systems [38].

In terms of *"ease of steering"*, designers mentioned the preference for *Text condition* over *Multi-modal condition* might be related to the learning curve and time constraints, but also due to their unfamiliarity with AI prompting and interfaces in general. Some novice users found the *Text condition* more straightforward to understand and control possibly because the interaction feels similar to traditional "type and search" mental models. Thus even without previous knowledge and any prompting strategies, users could adapt to text prompting input and the interface rather quickly, giving them more sense of control.

Controllability concerns also rose from the *automatic input interpretation* in *DesignPrompt*, as AI interpretations of images into semantic labels could cause loss of granularity and errors. Using semantic labels as direct prompts for tuning image input might be inspiring and efficient for some designers, but also comes with risks of concessions of user agency and failing users' expectations of AI's descriptive capabilities. This may let users desire a more straightforward way of control rather than choosing an expressive way of interaction. When designing *DesignPrompt*, we encountered similar trade-offs. Users depend on system interpretation to enable specific actions, such as utilizing semantics recognized for searching. However, they also have reservations about excessive system interpretation as they desire their intentions to be accurately conveyed. Designing for human-AI collaboration, however, requires a balance between these two imperatives.

Reflecting and expanding on RQ1, designers who are comfortable and confident with multimodal input can express themselves more effectively. Users reflected on the interactive prompt construction, and mentioned "prompt revealing" made them think the output was more "predictable" (RQ2). However, AI's misinterpretations of input material, users' lack of knowledge of AI or perceiving systems as not "easy to steer" makes designers feel "out of control" which results in expectation misalignment.

9.3 Perceptions of "Usefulness"

Some designers find AI-generated images novel and inspiring, while those seeking specific results deem them less useful. Some designers kept exploring visual inspirations from both search and GenAI throughout the *Multi-modal condition*. They used GenAI rather as an assistant helping them quickly experiment and combine inspiration. Others who relied more on searched images mainly wanted to modify small parts rather than generate something novel, also had higher expectations on precision and predictability of the GenAI outcome. Designers felt the *Multi-modal condition* encouraged them to have more abstract and less precise prompts, which seemed to be a good fit in the divergent ideation phase. However, the less precise prompts can also give excessive creative freedom to AI, generating illogical or inconsistent images. This may explain why in the questionnaire, only 25% of designers think the image generated from *Multi-modal condition* "makes more sense" and "more useful" compared with *Text condition*. Hence designers in a more divergent phase of their design process might have more open expectations where image usefulness becomes less defined, while in a convergent design phase, designers' expectations become more defined, impacting the perceived usefulness of the generated image.

Recent work on prompting for non-AI experts [70] raises the question of how tools can effectively establish capability expectations for end users. If users' expectations align with the system's capabilities, CSTs like *DesignPrompt* could provide users with greater transparency regarding their strengths and weaknesses. This would assist users in developing a more holistic design workflow.

Extended on discussions of RQ2, we see designers' perceived alignments between expectations and generated results can also depend on the divergent-convergent design processes they are in.

9.4 System Feature Appropriation

During the *Multi-modal condition* designers learned about the system and appropriated it in unexpected ways. For some designers, familiarizing themselves with the system shaped their creative process and perceptions of the tool. Regarding the use of interactive prompts, some users found the system more "transparent" after using it and also learned more about how prompts impact AI overall.

We noticed two interesting system usages in *Multi-modal condition*. One is ad-hoc exploration of a big cluster of visual inspirations: some designers have constructed extremely long visual prompts that would take much more effort to write as text. The other usage is inspiring element fixation: One designer used the same elements and colors in the *Multi-modal condition* input as a static pattern throughout the generation process, which allows to style generated image more consistently while still keeping ideas novel.

Feature appropriation directly relates to our RQ1, where multi-modalities not only enable users to express and explore better, but also allow them to uncover and explore unusual usages of multi-modalities. This also contributes to answer RQ3, as revising GenAI input interactively lets users perceive the system as more transparent, and by interacting with the system and prompts users can learn about the system and find it more useful over time.

9.5 Limitations

The structured observation study is designed to compare two variations of *DesignPrompt*, one with multimodal interactive prompt interface, the other with text prompt interface as baseline. We aimed to compare *DesignPrompt* with a state-of-art system. Our baseline system (with moodboard tool, search engine, text prompting and inpainting GenAI interface) is beyond existing standards and not available yet in practice which might have lowered the impact of our study results. However, since we aim to investigate the impacts caused by only interaction differences between *DesignPrompt* and other moodboard tools, the two variants we have chosen allowed us to make a fair comparison. In addition, 12 minutes for a moodboarding task is relatively short, as moodboard creation can take hours to days in practice, making our controlled task a compressed version of realistic moodboarding, especially given that the GenAI tool is still relatively new. More time would be required in future studies for a more extensive evaluation of *DesignPrompt*.

DesignPrompt uses semantic labels as "translations" from tuning images and colors to textual prompts. From an engineering perspective, it appears that using more refined textual inputs, such as incorporating semantic labels into general prompt templates or creating a comprehensive prompt using LLMs, might have a great potential to produce high-quality image outputs. However, more refined prompts using LLMs come at the cost of sacrificing user control and transparency. The direct mapping between multimodal input and "machine translations" in *DesignPrompt* allows users to have the flexibility to rearrange and modify the text prompts generated by the machine in interactive prompt editing. We see such "translation" as a design decision to be made.

9.6 Future Work

In the context of moodboard design, *DesignPrompt* exemplifies applying multimodal input in GenAI that facilitates expression and exploration in visual ideation. Our results show the overall benefits of visual-oriented multimodal input. However, it also adds another layer of understanding of interaction compared with text prompts. Further studies on investigating different prompt strategies' are necessary to explore the potential interaction space. This could include, for example, identifying stylistic and specific visual properties as important factors that influence users' image preferences, which are difficult to capture and extract in a textual context. We also see opportunities to add GenAI image editing techniques such as DreamBooth [54], ControlNet [71] and InstructPix2Pix [8] to *DesignPrompt* to allow designerly understandable interaction while designers can control visual details more expressively.

In addition to moodboards, the multimodal and interactive prompt pattern in *DesignPrompt* may be further useful in a broader context. It can be applied in a variety of design ideation methods, including storyboarding, brainstorming, and mindmapping, as well as more abstract creative processes like concept sketching and design iterations. Overall, we hope that the design implications identified during the pre-study will contribute to future GenAI-powered design CSTs.

10 CONCLUSION

We present *DesignPrompt*, a GenAI-powered digital moodboard tool that uses multimodal input including images, colors and semantics to help designers explore and express their intentions better to AI. Our pre-study has informed the design and implementation of *DesignPrompt* with design implications on translating users' intents to richer prompts.

The multimodal input interface of *DesignPrompt* provides designers opportunities to use different inspiration modalities to convey their ideas visually and abstractly. With flexibility in input combinations, *DesignPrompt* allows designers to fix certain inspiring elements and explore the middle ground between consistent themes and novel inspirations, while also keeping the options of pure literal and visual generation open. The interactions and layout in *DesignPrompt* are designed to be consistently uniform, which allows designers to use semantics, colors and images across image information, canvas and different sections in the AI generation track. *DesignPrompt* also encourages designers to learn and appropriate different features innovatively.

Comparing *DesignPrompt* with a realistic baseline, our structured observation study has investigated the expressive and exploratory power of multimodal input; the impacts of diverse interpretations of "feeling in control" on users' preferred ways of GenAI interaction; designers' perceptions of GenAI "usefulness" depending on their current design phases; and users' learning and appropriation with the system.

DesignPrompt demonstrates how we can empower design practice with GenAI using more human-centered design approaches. Apart from *DesignPrompt* system, our research also identified empirical design implications for future research directions.

ACKNOWLEDGMENTS

The presented work was funded by the European Commission as part of the Humane-AI-net project, under grant agreement number 952026. The work was further supported by Université Paris-Saclay's doctoral program funding (D23-ET046).

REFERENCES

[1] Adobe. 2023. Adobe Firefly. https://firefly.adobe.com/. [Accessed 08-02-2024].

[2] Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. 2023. MusicLM: Generating Music From Text. arXiv:2301.11325 [cs.SD]

[3] Jim Torresen Benedikte Wallace, Kristian Nymoen and Charles Patrick Martin. 2024. Breaking from realism: exploring the potential of glitch in AI-generated dance. *Digital Creativity* 0, 0 (2024), 1–18. https://doi.org/10.1080/14626268.2024.2327006 arXiv:https://doi.org/10.1080/14626268.2024.2327006

[4] Adrien Bousseau, Theophanis Tsandilas, Lora Oehlberg, and Wendy E. Mackay. 2016. How Novices Sketch and Prototype Hand-Fabricated Objects. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 397–408. https://doi.org/10.1145/2858036.2858159

[5] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3586183.3606725

[6] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[7] Andrew Brock, Jeff Donahue, and Karen Simonyan. 2019. Large Scale GAN Training for High Fidelity Natural Image Synthesis. arXiv:1809.11096 [cs.LG]

[8] Tim Brooks, Aleksander Holynski, and Alexei Efros. 2023. InstructPix2Pix: Learning to Follow Image Editing Instructions. 18392–18402. https://doi.org/10.1109/CVPR52729.2023.01764

[9] Daniel Buschek, Lukas Mecke, Florian Lehmann, and Hai Dang. 2021. Nine Potential Pitfalls when Designing Human-AI Co-Creative Systems. *CoRR* abs/2104.00358 (2021). arXiv:2104.00358 https://arxiv.org/abs/2104.00358

[10] Tracy Cassidy. 2008. Mood boards: Current practice in learning and teaching strategies and students' understanding of the process. *International Journal of Fashion Design* Technology and Education (03 2008), 43–54. https://doi.org/10.1080/17543260802015154

[11] Li-Yuan Chiou, Peng-Kai Hung, Rung-Huei Liang, and Chun-Teng Wang. 2023. Designing with AI: An Exploration of Co-Ideation with Image Generators. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1941–1954. https://doi.org/10.1145/3563657.3596001

[12] John Joon Young Chung. 2022. Artistic User Expressions in AI-powered Creativity Support Tools. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 95, 4 pages. https://doi.org/10.1145/3526114.3558531

[13] John Joon Young Chung and Eytan Adar. 2023. PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3586183.3606777

[14] John Joon Young Chung, Shiqing He, and Eytan Adar. 2021. The Intersection of Users, Roles, Interactions, and Technologies in Creativity Support Tools. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1817–1833. https://doi.org/10.1145/3461778.3462050

[15] Comfyanonymous. [n. d.]. GitHub - comfyanonymous/ComfyUI: The most powerful and modular stable diffusion GUI, api and backend with a graph/nodes interface. https://github.com/comfyanonymous/ComfyUI

[16] Niklas Deckers, Maik Fröbe, Johannes Kiesel, Gianluca Pandolfo, Christopher Schröder, Benno Stein, and Martin Potthast. 2023. The Infinite Index: Information Retrieval on Generative Text-To-Image Models. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 172–186. https://doi.org/10.1145/3576840.3578327

[17] Prafulla Dhariwal and Alex Nichol. 2021. Diffusion Models Beat GANs on Image Synthesis. arXiv:2105.05233 [cs.LG]

[18] Kees Dorst and Nigel Cross. 2001. Creativity in the design process: co-evolution of problem-solution. *Design Studies* 22, 5 (2001), 425 – 437. http://www.sciencedirect.com/science/article/pii/S0142694X01000096

[19] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 278–288. https://doi.org/10.1145/3025453.3025739

[20] 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 (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300619

[21] Jérémie Garcia, Theophanis Tsandilas, Carlos Agon, and Wendy E. Mackay. 2014. Structured observation with polyphony: a multifaceted tool for studying music composition. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (Vancouver, BC, Canada) *(DIS '14)*. Association for Computing Machinery, New York, NY, USA, 199–208. https://doi.org/10.1145/2598510.2598512

[22] William Gaver. 2011. Making spaces: how design workbooks work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1551–1560. https://doi.org/10.1145/1978942.1979169

[23] GoogleCloud. 2023. Vision AI|cloud vision API|google cloud. https://cloud.google.com/vision. [Online; accessed 3-June-2023].

[24] Jonathan Ho, Chitwan Saharia, William Chan, David J. Fleet, Mohammad Norouzi, and Tim Salimans. 2022. Cascaded Diffusion Models for High Fidelity Image Generation. *Journal of Machine Learning Research* 23, 47 (2022), 1–33.

[25] Nisha Huang, Fan Tang, Weiming Dong, and Changsheng Xu. 2022. Draw Your Art Dream: Diverse Digital Art Synthesis with Multimodal Guided Diffusion. In *Proceedings of the 30th ACM International Conference on Multimedia* (, Lisboa, Portugal,) *(MM '22)*. Association for Computing Machinery, New York, NY, USA, 1085–1094. https://doi.org/10.1145/3503161.3548282

[26] Angel Hsing-Chi Hwang. 2022. Too Late to be Creative? AI-Empowered Tools in Creative Processes. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 38, 9 pages. https://doi.org/10.1145/3491101.3503549

[27] Ali Jahanian, Jerry Liu, Qian Lin, Daniel Tretter, Eamonn O'Brien-Strain, Seungyon Claire Lee, Nic Lyons, and Jan Allebach. 2013. Recommendation system for automatic design of magazine covers. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces* (Santa Monica, California, USA) *(IUI '13)*. Association for Computing Machinery, New York, NY, USA, 95–106. https://doi.org/10.1145/2449396.2449411

[28] 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 (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3411764.3445093

[29] Youwen Kang, Zhida Sun, Sitong Wang, Zeyu Huang, Ziming Wu, and Xiaojuan Ma. 2021. MetaMap: Supporting Visual Metaphor Ideation through Multi-dimensional Example-based Exploration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (, Yokohama, Japan,) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 427, 15 pages. https://doi.org/10.1145/3411764.3445325

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

[31] Taslim Arefin Khan, Dongwook Yoon, and Joanna McGrenere. 2020. *Designing an Eyes-Reduced Document Skimming App for Situational Impairments*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376641

[32] 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 (IUI '23)*. Association for Computing Machinery, New York, NY, USA, 919–933. https://doi.org/10.1145/3581641.3584078

[33] Janin Koch. 2020. *Collaborative Systems for Design Inspiration*. Aalto University. https://aaltodoc.aalto.fi:443/handle/123456789/46910 Accepted: 2020-10-13T09:00:05Z ISSN: 1799-4942 (electronic).

[34] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300863

[35] Janin Koch, Nicolas Taffin, Michel Beaudouin-Lafon, Markku Laine, Andrés Lucero, and Wendy E. Mackay. 2020. ImageSense: An Intelligent Collaborative Ideation Tool to Support Diverse Human-Computer Partnerships. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (May 2020), 45:1–45:27. https://doi.org/10.1145/3392850

[36] Janin Koch, Nicolas Taffin, Andrés Lucero, and Wendy E. Mackay. 2020. SemanticCollage: Enriching Digital Mood Board Design with Semantic Labels. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 407–418. https://doi.org/10.1145/3357236.3395494

[37] Kosinkadink. [n. d.]. GitHub - Kosinkadink/ComfyUI-Advanced-ControlNet: ControlNet scheduling and masking nodes with sliding context support. https://github.com/Kosinkadink/ComfyUI-Advanced-ControlNet

[38] Q. Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly Understanding: Information Needs for Model Transparency to Support Design Ideation for AI-Powered User Experience. http://arxiv.org/abs/2302.10395 arXiv:2302.10395 [cs].

[39] Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. 2021. What Makes Good In-Context Examples for GPT-3? arXiv:2101.06804 [cs.CL]

[40] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3491102.3501825

[41] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating Text-to-Image AI in 3D Design Workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (, Pittsburgh, PA, USA,) *(DIS

'23). Association for Computing Machinery, New York, NY, USA, 1955–1977. https://doi.org/10.1145/3563657.3596098

[42] Lokesh Dhakar. 2023. COLOR THIEF. https://lokeshdhakar.com/projects/color-thief/. [Online; accessed 5-June-2023].

[43] 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 *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3491101.3519809

[44] Andrés Lucero. 2012. Framing, aligning, paradoxing, abstracting, and directing: how design mood boards work. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. Association for Computing Machinery, New York, NY, USA, 438–447. https://doi.org/10.1145/2317956.2318021

[45] Wendy E. Mackay and Joanna McGrenere. to appear in 2024. Comparative Structured Observation. *ACM Transactions on Computer-Human Interaction (TOCHI)*.

[46] Mehta, Chirag. 2007. Name that color. https://chir.ag/projects/name-that-color/. [Online; accessed 5-June-2023].

[47] Midjourney. 2024. Midjourney. https://www.midjourney.com/. [Accessed 26-01-2024].

[48] Emad Mostaque. 2023. Stable Diffusion Public Release — Stability AI — stability.ai. https://stability.ai/news/stable-diffusion-public-release. [Accessed 26-01-2024].

[49] Mohammad Amin Mozaffari, Xinyuan Zhang, Jinghui Cheng, and Jin L.C. Guo. 2022. GANSpiration: Balancing Targeted and Serendipitous Inspiration in User Interface Design with Style-Based Generative Adversarial Network. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3491102.3517511

[50] OpenAI. 2023. OpenAI platform. https://platform.openai.com/docs/api-reference. [Online; accessed 5-June-2023].

[51] Openai. 2024. DALL·E 3 — openai.com. https://openai.com/dall-e-3. [Accessed 26-01-2024].

[52] Jonas Oppenlaender. 2022. Prompt engineering for text-based generative art. *arXiv preprint arXiv:2204.13988* (2022).

[53] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. 2022. DreamFusion: Text-to-3D using 2D Diffusion. arXiv:2209.14988 [cs.CV]

[54] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2023. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[55] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. arXiv:2205.11487 [cs.CV]

[56] Téo Sanchez. 2023. Examining the Text-to-Image Community of Practice: Why and How Do People Prompt Generative AIs?. In *Proceedings of the 15th Conference on Creativity and Cognition (C&C '23)*. Association for Computing Machinery, New York, NY, USA, 43–61. https://doi.org/10.1145/3591196.3593051

[57] Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M. Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal V. Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Févry, Jason Alan Fries, Ryan Teehan, Stella Biderman, Leo Gao, Tali Bers, Thomas Wolf, and Alexander M. Rush. 2021. Multitask Prompted Training Enables Zero-Shot Task Generalization. *CoRR* abs/2110.08207 (2021). arXiv:2110.08207 https://arxiv.org/abs/2110.08207

[58] Yang Shi, Yang Wang, Ye Qi, John Chen, Xiaoyao Xu, and Kwan-Liu Ma. 2017. IdeaWall: Improving Creative Collaboration through Combinatorial Visual Stimuli. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 594–603. https://doi.org/10.1145/2998181.2998208

[59] Patrick C. Shih, David H. Nguyen, Sen H. Hirano, David F. Redmiles, and Gillian R. Hayes. 2009. GroupMind: supporting idea generation through a collaborative mind-mapping tool. In *Proceedings of the 2009 ACM International Conference on Supporting Group Work* (, Sanibel Island, Florida, USA,) *(GROUP '09)*. Association for Computing Machinery, New York, NY, USA, 139–148. https://doi.org/10.1145/1531674.1531696

[60] Hariharan Subramonyam, Roy Pea, Christopher Lawrence Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the Gulf of Envisioning: Cognitive Design Challenges in LLM Interfaces. arXiv:2309.14459 [cs.HC]

[61] Lucille Alice Suchman. 2007. *Human-machine reconfigurations: Plans and situated actions.* Cambridge university press.

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

[63] Unsplash. [n. d.]. Unsplash Image API | Free HD Photo API — api.unsplash.com. https://api.unsplash.com/. [Accessed 29-01-2024].

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

[65] Yael Vinker, Andrey Voynov, Daniel Cohen-Or, and Ariel Shamir. 2023. Concept Decomposition for Visual Exploration and Inspiration. *ACM Trans. Graph.* 42, 6, Article 241 (dec 2023), 13 pages. https://doi.org/10.1145/3618315

[66] Hao-Chuan Wang, Dan Cosley, and Susan R. Fussell. 2010. Idea expander: supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 103–106. https://doi.org/10.1145/1718918.1718938

[67] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv:2302.11382 [cs.SE]

[68] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376301

[69] Nur Yildirim, Changhoon Oh, Deniz Sayar, Kayla Brand, Supritha Challa, Violet Turri, Nina Crosby Walton, Anna Elise Wong, Jodi Forlizzi, James McCann, and John Zimmerman. 2023. Creating Design Resources to Scaffold the Ideation of AI Concepts. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2326–2346. https://doi.org/10.1145/3563657.3596058

[70] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3544548.3581388

[71] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 3813–3824. https://doi.org/10.1109/ICCV51070.2023.00355