Exploring Visual Prompts: Refining Images with Scribbles and Annotations in Generative AI Image Tools

Hyerim Park
BMW Group
Munich, Germany
University of Stuttgart
Stuttgart, Germany
hyerim.park@bmw.de

Malin Eiband
BMW Group
Munich, Germany
malin.eiband@bmw.de

Andre Luckow
BMW Group
Munich, Germany
LMU Munich
Munich, Germany
andre.luckow@bmwgroup.com

Michael Sedlmair
University of Stuttgart
Stuttgart, Germany
michael.sedlmair@visus.uni-stuttgart.de

Abstract

Generative AI (GenAI) tools are increasingly integrated into design workflows. While text prompts remain the primary input method for GenAI image tools, designers often struggle to craft effective ones. Moreover, research has primarily focused on input methods for ideation, with limited attention to refinement tasks. This study explores designers' preferences for three input methods—text prompts, annotations, and scribbles—through a preliminary digital paper-based study with seven professional designers. Designers preferred annotations for spatial adjustments and referencing in-image elements, while scribbles were favored for specifying attributes such as shape, size, and position, often combined with other methods. Text prompts excelled at providing detailed descriptions or when designers sought greater GenAI creativity. However, designers expressed concerns about AI misinterpreting annotations and scribbles and the effort needed to create effective text prompts. These insights inform GenAI interface design to better support refinement tasks, align with workflows, and enhance communication with AI systems.

CCS Concepts

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

Keywords

generative AI, text prompts, annotation-based input, scribble-based input, design refinement

1 Introduction and Background

Generative AI (GenAI) image generation tools[1] enable users to create high-quality images from simple inputs like text and images, producing outputs comparable to human-created designs. Popular tools include Midjourney [20], DALL-E 3 (integrated into ChatGPT) [1], Adobe Firefly [10], and DreamStudio (built on Stable Diffusion) [8]. GenAI has been widely adopted across fields such as industrial

---

[1]In this paper, 'GenAI' refers specifically to generative AI for image generation.

design [11, 19], graphic design [5, 14, 28, 30], illustration [3, 18], UI/UX design [9, 15, 24], and fashion design [31]. With its growing adoption, designing effective interfaces has become a central research focus in Human-Computer Interaction (HCI), particularly in supporting professional designers [5, 6, 23]. However, qualitative studies highlight that designers often struggle with crafting text prompts, which are currently the dominant input method [4, 11, 22]. Mismatches between intended designs and generated outputs frequently require corrections, disrupting workflows and diverting attention from core design tasks [26, 33]. GenAI's limited contextual understanding and reliance on trial-and-error further complicate iterative workflows, frequently requiring designers to switch tools and increasing cognitive load [22, 27].

To address these challenges, researchers have suggested enhancing text-prompt interfaces with features such as automated keyword suggestions [5, 23], revised prompt generation [3, 25], and dropdown menus for interactive refinement [29]. In addition to improving text-based inputs, multimodal approaches incorporating visual inputs have also been explored. Examples include combining text with image uploads, color palettes [23], or rough sketches [7], as well as introducing visual controls like sliders for blending prompt weights or anchors for resizing elements [6]. However, these visual interactions primarily supplement text prompts rather than serve as primary input methods.

Research on GenAI image tools has mostly focused on ideation and exploration, where these tools support divergent thinking by generating various concepts and inspiring creativity [3–5, 17, 25, 27]. However, refinement tasks, which involve making precise adjustments to better align with user intent, remain underexplored. Our study defines *refinement* as an iterative process of generating and fine-tuning images using GenAI to match a designer's intended vision, which may evolve through interaction with AI-generated outputs. In contrast, *finalization* focuses on meticulous, pixel-perfect modifications aimed at delivering a polished, production-ready result.

Several studies highlight the need for better refinement support in GenAI tools. Users often seek iterative, localized regeneration to modify specific elements [33]. As designs evolve, the focus shifts

from broad characteristics like style and layout to finer details and spatial composition [7]. While GenAI tools typically prioritize ideation, refinement tasks demand higher precision and iterative capabilities, which many existing tools lack [17]. Current *inpainting* techniques enable users to refine specific image regions by marking areas with brushes, erasers, or stencils, seamlessly blending new content with surrounding elements [12, 13, 32]. However, these methods predominantly rely on text prompts to guide edits, which designers often find limiting [7, 17, 22].

Scribbles and annotations are widely used in design workflows and have potential for refinement tasks. Designers often begin with scribbles, rough sketches, layout drawings, or handwritten annotations to visualize their ideas before working with GenAI [7]. When ideas are vague or underdeveloped, some designers find it challenging to articulate them through text prompts [22, 27], increasing interest in integrating visual inputs with GenAI tools [22]. Some research has differentiated between text-based and visual-based inputs, highlighting their distinct roles in the creative process. For example, one study suggests that sketch-based inputs are more effective for refining and embodying designs in later stages, while text prompts are generally better suited for early ideation [16]. Similarly, designers often transition from text inputs to image-based inputs for localized variations and then switch back to text prompts to explore new directions [25]. To investigate these dynamics, we examine how scribbles and annotations compare to text inputs in refinement tasks using GenAI. Our research is guided by the following question:

> **RQ:** How do professional designers perceive and prefer annotations, scribbles, and text-based input methods for different refinement tasks in GenAI image tools?

We conducted a preliminary user study using a digital paper-based prototype with seven professional designers specializing in UI/UX, automotive interior, graphics, and XR design. Participants completed five types of refinement tasks using text, annotation, and scribble inputs. Our findings revealed varied preferences among input methods. Annotations were preferred for spatial adjustments (e.g., moving, resizing) and referencing in-image elements (e.g., applying attributes from one object to another). Scribbles were used to specify attributes like shape, position, and size, often complemented by other methods for additional detail. Text inputs were favored for lengthy descriptions involving multiple requirements or tasks that relied on more significant GenAI influence. While experienced users expressed more confidence in text prompts, some participants raised concerns about AI's ability to interpret rough scribbles, handwriting, or custom visual symbols (e.g., arrows, numbering, multi-colored annotations). These findings highlight the potential of diverse input methods beyond text prompts to support refinement tasks in GenAI image tools, encouraging further exploration to enhance them for later stages of the creative process.

2 User Study

We conducted a preliminary digital paper-based user study to explore designers' preferences for different input methods in refinement tasks using GenAI image tools. Participants interacted with three input methods (see Figure 1):

- **Text Prompts:** Typed textual instructions entered via a keyboard.
- **Annotations:** Text or visual symbols (e.g., arrows, circles) added directly to the image using a stylus or mouse.
- **Scribbles:** Freeform sketches drawn directly on the image using a stylus or mouse.

Marking Edit Areas: All three methods employed the *inpainting* technique, using a selection brush to mark areas for editing, following standard practices in commercial tools such as Midjourney, Adobe Firefly, and DALL-E [10, 12, 13]. Based on feedback from our pre-tests (N=2), we designed annotations and scribbles to allow users to mark areas directly (e.g., circles, arrows) using the same pen tool, avoiding additional menu navigation. In contrast, text prompts, typed with a keyboard, typically require a selection tool to specify the inpainting area before applying edits. In our user study, most participants used the pen tool for annotations and scribbles rather than the inpainting selection brush.

2.1 Participants

Seven designers from an international automotive company participated in the study, recruited via outreach and snowball sampling. Their roles covered UI/UX, graphic, interior, and XR design, with projects extending beyond automotive into broader design contexts. GenAI usage frequency varied: daily (1), 2–3 times weekly (2), weekly (2), monthly (1), and less than monthly (1). Participants primarily used Midjourney, DALL-E 3, and Adobe Firefly. Each in-person session lasted approximately 70 minutes.

2.2 Procedure

The study consisted of four phases: (1) introduction and tutorials, (2) pre-tasks, (3) main tasks, and (4) interviews. A think-aloud method was employed, with participants' verbalizations recorded and supplemented by the experimenter's observation notes.

(1) Introduction and Tutorials: Participants were briefed on the study's objectives, procedures, and consent process. A tutorial introduced DALL-E 3's text prompting and inpainting capabilities, reflecting the primary methods in most GenAI tools. The three input methods—text prompts, annotations, and scribbles—were explained within the context of our simulated settings. Input methods were tested without generating actual outputs to minimize bias from system capabilities (e.g., output quality, latency). Participants selected their preferred input method for each task, with the option to use multiple methods if they favored them equally. For the hardware setup, they could choose between a tablet with a stylus or a laptop with a mouse. During the main tasks, we observed that all participants used a stylus for scribbles and annotations.

(2) Pre-Tasks: Participants completed two pre-tasks to familiarize themselves with the input methods: **(PT1)** *adding a simple Santa hat to a teddy bear* and **(PT2)** *modifying the Santa hat into a highly detailed crown.* For each task, they chose their preferred input method(s) and explained how they would guide GenAI to complete it.

(3) Main Tasks: Participants generated several images using DALL-E 3 for presentations, team event posters, or other non-confidential creative projects. After selecting one favorite image, they completed six refinement tasks: **(T1)** *adding new elements,*

Text Prompts ⌨

Inpainting

Add a red bow.

Annotations ✏

Scribbles 🖌

Figure 1: Illustration of the Three Input Methods Tested in the Study: (from left to right) text prompts (typed instructions requiring *inpainting* **for area selection), annotations (text or visual symbols on the image, with optional** *inpainting* **selected by the user), and scribbles (freeform scribbles on the image, with optional** *inpainting* **selected by the user).**

(T2) *modifying specific elements (e.g., enhancing complexity),* **(T3)** *revising materials or surfaces,* **(T4)** *making structural or spatial adjustments,* **(T5)** *performing global modifications,* and **(T6)** *a free task to revise any aspect of their choice.* The task order was randomized using a Latin square design, with specific task details adapted to each participant's generated images. While the experimenter suggested refinement tasks, participants were encouraged to propose modifications within the task categories to better align with their design goals and enhance their motivation.

(4) Interviews: Semi-structured interviews followed the tasks, exploring participants' experiences with each input method. Discussions covered their advantages, limitations, task-specific preferences, alignment with real-world workflows, and professional challenges. Participants also provided suggestions for improving each input method.

3 Findings

An iterative review of participant feedback and observations revealed patterns in input preferences (see Table 1). Although the study introduced three distinct input methods, participants frequently combined them in various ways. For instance, scribbles were often used alongside annotations and text prompts to add details and improve clarity during refinement tasks. While our study primarily focused on refinement, participant feedback also provided insights into the ideation process. We observed that refinement and ideation were not always distinct, as participants frequently shifted between them. Moreover, input methods were not strictly tied to a single stage of the creative process. Participants selected different methods based on the task at hand or the clarity of their design intent. We include these observations to offer a more comprehensive understanding of input preferences.

3.1 Text Prompts: Effective for Detailed Descriptions and Global Refinements, but Challenging to Craft

Detailed Instructions. When designers had a **clear intent for modification** and a confident **grasp of how to describe it**, they

preferred text inputs for tasks requiring **lengthy and detailed descriptions** to guide GenAI, especially when specifying **multiple attributes of a target element**. For example, participants provided prompts such as *"Add a Santa hat in wine-red color, made of velvet, and with sequin decoration"* (P6) or *"Replace her dress with a design that looks like a collaboration with Louis Vuitton designers, incorporating a more luxurious concept"* (P7).

Global Modifications. Text inputs proved effective for tasks involving **global changes**, such as modifying an image's **overall tone or format**. Examples included prompts like *"Remove the yellow tone"* (P6) and *"Change the image ratio to be more square"* (P4). The brush width control in *inpainting* allowed for a more efficient selection of large areas compared to manual annotations (e.g., circling elements) when selecting substantial parts of elements in the image (P5, P6, P7). Designers often used text prompts to specify **styles**, referencing **brands, designer names, or design concepts** to achieve desired **moods or aesthetics** (P6, P7).

Creative Flexibility. Text inputs were also preferred when participants had **uncertainty about their design intent or concept** or wanted to leverage **GenAI's creative input**. Examples included replacing a male figure with a generic female figure for *"a girlfriend look"* (P5), refining kiosk textures to a *"bolder style"* (P3), or adjusting a menu display to a *"more minimal style"* (P1). P3 mentioned relying on AI due to having *"no ideas about kiosk surfaces,"* while P1 appreciated AI's ability to *"find additional details."*

Familiarity Among Experienced Users. Three designers who regularly used GenAI tools (either daily or 2 to 3 times per week) preferred text prompts more often across tasks in the study. They credited this preference to **company training** and proactive efforts to **improve their prompt-crafting skills through online communities and resources**. P7 reflected that their preference might have differed without **long-term experience**, while P1 remarked, *"I am more familiar with text prompts."*

Challenges in Crafting Prompts. While effective, **crafting text prompts remains challenging**. P1 remarked, *"(…) explaining visual ideas in words is not easy."* Others (P2–P6) questioned GenAI's

Table 1: Participants' Input Method Preferences Across Tasks: Multiple methods in a single cell indicate an equal preference for those inputs. Methods in parentheses represent selections for the free task (T6), in which participants chose what to refine.

PID	(T1) Add New Elements	(T2) Modify Specific Elements	(T3) Revise Materials or Surfaces	(T4) Structural or Spatial Adjustments	(T5) Global Modifications
P1	Text prompts	Scribbles + annotations	Annotations, (T6: Scribbles + annotations)	Scribbles + annotations	Text prompts
P2	Scribbles + annotations	Annotations, (T6: Annotations)	Text prompts	Annotations, Scribbles + annotations	Text prompts
P3	Scribbles + annotations, Annotations	Annotations, (T6: Annotations)	Text prompts	Scribbles + annotations	Text prompts
P4	Scribbles + text prompts	Annotations	Text prompts	Annotations	Text prompts + annotations, (T6: Text prompts)
P5	Scribbles + text prompts, Scribbles + text prompts + annotations	Annotations	Text prompts	Annotations, (T6: Annotations)	Text prompts
P6	Text prompts	Scribbles + annotations, (T6: Text prompts)	Text prompts	Annotations, Text prompts	Text prompts
P7	Text prompts, (T6: Scribbles + text prompts)	Text prompts	Annotations	Annotations	Text prompts

ability to fully interpret their descriptions and noted that creating prompts required **significant time and effort**. Several participants (P3, P4, P6, P7) anticipated needing **multiple iterations** to achieve their desired outcomes in real-world scenarios. These challenges align with prior research highlighting the **cognitive effort** involved in effectively communicating design ideas [22, 26].

3.2 Annotations: Useful for Spatial and Attribute Referencing, but Concerns About GenAI Interpretation

Spatial Adjustments and In-Image Attribute Referencing. Annotations were commonly used for **spatial adjustments**, such as repositioning a plant or resizing lamps or logos. Designers often used circles and arrows along with brief handwritten notes indicating what GenAI should do (e.g., *"replace"* or *"switch"*) or specifying a target object (e.g., *"ribbon"* or *"dachshund"*). Annotations also effectively supported **referencing attributes within an image**. Designers used arrows to indicate that a texture from one person's clothing should be applied to another's outfit (P3) or to match the size of two beer glasses by labeling them as *"1"* and *"2"*, then adding a text prompt like *"Make 1 and 2 the same size"* (P5).

Directional Guidance. For tasks requiring **directional instructions**, such as lighting direction adjustments or orienting elements (e.g., *"eyes," "audience seats"*), annotations—particularly symbols like **dots and arrows**—were preferred over text prompts for their speed and ease of use. For instance, designers marked lighting directions with arrows, adding instructions like *"Move the sun direction following the arrow"* (P7) or *"Make the eyes look at this point (dot)"* (P5). They noted that they would rely solely on symbols if AI could interpret them correctly (P7). Lighting effect adjustments, such as **brightening or shadowing surfaces**, were clarified with annotations paired with scribbles, as seen in P1's use of rough strokes combined with annotations like *"brighter"* or *"darker."*

Minor Refinements in Later Stages. Annotations were preferred for **small-scale adjustments that required minimal text input** in later refinement stages. For example, P1 and P7 noted that annotations were adequate for minor tasks or a single change, such as *"a bit softer"* or *"more padding"*, due to their simplicity and clarity. In contrast, annotations were less practical in early refinement stages, which often involved more extensive changes, as capturing all details required lengthy descriptions. For instance, P7 considered annotating grass with multiple instructions, such as *"fluffy," "denser with more grass,"* and *"more lighting effects on the surface,"* too labor-intensive by hand, making shorter descriptions prone to vagueness. To add precision, some designers **combined annotations with scribbles** to specify **exact dimensions**, such as marking a dog's size, position, and orientation or the spacing between menu buttons.

Workflow Integration. P6 described **an envisioned workflow** that involved **iterative refinements** in the order of text prompts, annotations, and rough scribbles. They typically began with text prompts, like *"Apply a floral pattern to the pants,"* to generate images that allowed for broad AI-driven modifications. Annotations were then used to fine-tune details, marking areas with arrows or circles and adding instructions like *"smaller"* to adjust the pattern size. If further clarification was needed, they conveyed their intent by quickly scribbling the desired size and shape directly onto the image.

Visibility and AI Interpretation Challenges. Despite their advantages, annotations had limitations. Overlapping annotations on images were seen as disruptive, hindering visual clarity. Without zoom features, **annotations could quickly obscure the image**, making finer details harder to discern. Some designers (P3-P6) suggested a **canvas-style layout**, allowing annotations to be written on an empty canvas outside the image. Concerns about **AI's interpretation of handwriting** were also raised (P1, P6). While annotations and scribbles in personal workflows were often **informal and quick**, participants in the study **made an effort to write clearly**, knowing their inputs were shared, even though it was not required.

3.3 Scribbles: Facilitating Simple Edits, but Requiring Complementary Input for Clarity

Precise Goals for Edits Focusing on a Single Attribute. Scribbles enabled users to focus on a **single attribute or a small set of attributes with precision**, especially when **combined with text prompts or annotations**. They were effective for indicating details such as the **placement of elements** (e.g., *"a ribbon on the dress"*) (P6), **shape, size, and angle** (e.g., *of "beanies"*) (P5), snow **distribution** on a mountain (P2), and **the number and position of players** (P5). Scribbles were also used to resize objects like lamps, dress lengths, and balloons to **specific dimensions** (P2, P3, P6, P7). This approach was particularly useful when users had **clear goals and a concrete idea** for modifying specific elements, allowing **precise marking** to indicate adjustments. For example, P1 employed scribbles to adjust the size and spacing of polka dots on fabric more effectively than using text prompts. Similarly, P4 noted, *"It is easier to draw than explain in text"* and sketched a station design directly on the image, helping expedite the iterative process while providing more control.

Universally Recognizable Concepts with Scribbles. Scribbles were preferred for **simple, universally recognized objects** such as *"polka dots," "not very detailed patterns"* (P1), or *"football, bowtie, or simple hat"* (P5). For more complex and unique elements, like a football team logo, users often combined scribbles with additional textual instructions, such as *"Add a Tottenham logo on the pink circle."* As P5 noted, *"It is harder to scribble complex designs as it takes time, and I doubt AI understands them."*

Combining with Other Inputs for Clarity. While scribbles offered advantages, most users avoided relying on them alone due to concerns about their **limited detail** and **uncertainty about AI's interpretation**. Scribbles were often combined with text prompts or annotations to provide context or additional details. For example, P7 said, *"Make the lamp bigger, as I drew it,"* while P4 added, *"Add a flying car station to match my scribble."* Scribbles also had limitations, particularly for expressing concepts like **materials and textures** (e.g., *"shiny surfaces"*) (P5), **abstract ideas** (e.g., *"minimal display"*) (P1), or **complex instructions that were difficult to convey through scribbles alone** (e.g., *"Make the building fit inside the image"*) (P6). Poorly drawn scribbles could mislead the AI, resulting in unintended outcomes. For example, P6 noted that rough sketches, such as a man with an exaggerated head and narrow shoulders, could lead to **distorted designs**. These concerns highlight the need to **combine scribbles with text descriptions** to improve clarity and reduce ambiguity.

Playful Exploration and Non-Designer Use. Some designers (P1, P6, P7) viewed scribbles as **better suited for non-designers**, enabling experimentation and playful exploration. P7 observed that scribbles work well for **general users** when precision in **elements like material, texture, color, or size is not critical.** However, P7 emphasized that professional design work necessitates **consistent visual quality**.

Beneficial for both Ideation and Final Adjustments. Some designers found scribbles **useful for ideation tasks**, especially when instructing **rough outlines of elements** in an image. Scribbles allowed quick iterations and reduced generation time without requiring precise descriptions of layout and composition. P3 envisioned combining very rough sketches, resembling blocks, with text annotations **to guide the position, size, and layout of elements in initial images.** Conversely, scribbles were also valued in later stages, with P6 noting that they would be particularly useful for **fine marking** to make **precise adjustments** (e.g., subtle refinements in shape). This approach could save time by reducing the need to create text prompts for minor modifications or export work to design tools like Photoshop.

4 Key Insights and Future Research Directions

In this section, we present key insights from our findings and outline potential research questions and directions to further explore and enhance GenAI image tool interfaces for design workflows. Although our study focuses on GenAI in refinement tasks, some directions extend to the broader creative process. As a qualitative preliminary study, our insights also provide useful context for understanding designers' interactions with these tools across workflows.

4.1 How Can GenAI Dynamically Adapt Input Methods to Meet Diverse User Needs and Contexts?

Input preferences varied based on the user's need to balance clarity of edits with openness to AI's creative influence. For precise refinements, annotations or scribbles—often combined with other methods—were preferred for directly and efficiently specifying changes such as size, position, or shape. In contrast, text prompts were favored for open-ended tasks, offering greater room for AI interpretation and creativity, such as global adjustments to styles (e.g., *"minimal," "futuristic"*) or material changes (e.g., *"matte"*). The balance between control and flexibility shifts across creative phases, projects, and user expertise. Key questions include: How can GenAI adapt to diverse user needs and contexts, providing the proper input methods at the right time? Additionally, how can it enable seamless transitions between methods without disrupting workflows? Exploring these questions could guide the design of more intuitive, adaptable systems.

4.2 How Can GenAI Interfaces Adapt to Varied Work Setups?

All participants opted for styluses for annotations and scribbles when provided. Designers who primarily work with styluses were more inclined to envision using these inputs in GenAI, citing compatibility with their workflows. In contrast, those using a mouse or trackpad were less likely to favor these methods due to workflow disruptions, except for one designer who noted that GenAI's iterative nature, requiring multiple edits, made switching tools worthwhile for refinements. Input preferences may shift in smartphone environments that rely on finger interactions. While not tested, designer interviews and tools like Galaxy AI's sketch-to-image feature [2] suggest finger-based scribbles may appeal to non-designers,

enabling creativity with less precision. Future research could explore how input preferences vary across devices, highlighting user group differences and how GenAI input methods adapt to different setups. For instance, annotations could integrate with a keyboard and mouse through features like sticky notes (e.g., Figma, Miro) or comment tools in document applications (e.g., Word, Google Docs), broadening their applicability across different work setups.

4.3 How Can GenAI Interfaces Better Interpret User Intent in Annotations?

Regarding annotation input, participants raised concerns about AI misinterpreting handwriting or simple symbols like arrows, often supplementing them with text prompts for clarity (e.g., *"apply changes to the circle"*). While annotations are typically informal and self-explanatory, participants made extra efforts to ensure AI comprehension during the study. Key questions arise: How can GenAI better differentiate between annotation types (e.g., symbols, handwriting, labeling) to interpret user intent? What features, such as color coding or numbering, could help users convey instructions more clearly and reduce errors? Integrating conventional features like canvas layouts, zoom, or annotation notes for meta-communication could enhance annotation creation and management. Exploring common annotation symbols and conventions may also guide more intuitive, user-centered GenAI interface design.

4.4 Should GenAI Interfaces Adapt to Users or Guide Users to Adapt?

Our study found that designers who frequently used GenAI tools (daily or 2–3 times per week) and engaged in training and online communities showed a stronger preference for text prompts than others. While they acknowledged the challenges of crafting prompts, their skills and confidence improved significantly with experience, easing the initial learning curve. This observation raises a key question: Should systems accommodate users' existing skills, or should users adapt through training and iterative practice? This balance in current GenAI image tools could be explored in two directions [21]: enhancing support for crafting effective text prompts, encouraging adaptation to text-centered tools, or improving alternative input methods to better support visually oriented designers.

5 Limitations

This study has several limitations. First, the small sample size may limit the generalizability of our findings. While we identified patterns, a larger participant pool could reveal additional preferences and nuances. However, as a preliminary study, we believe it provides a valuable starting point for understanding input preferences for refinement. Second, the simulated input setting, which isolated input methods from factors like latency or output quality (which could influence trial-and-error cycles), meant that participants did not produce actual outputs. This may have affected their perceptions and limited the depth of their overall experience. Additionally, this preliminary study observed only a single instance of input creation for specific tasks. We acknowledge that this does not fully reflect the inherently iterative nature of refinement, where designers adjust their inputs after reviewing AI-generated results, influencing

their preferred input methods. Moreover, while we categorized refinement tasks into five main types and included a free task, all free tasks chosen by designers fell within our predefined categories. Additional task types could emerge through long-term real-world observations or larger studies. Lastly, while our findings highlight the distinct roles of annotations and scribbles, we also observed some overlap, as designers often used both together for tasks like spatial adjustments and referencing elements. Future research could further explore their unique use cases and how they can be effectively combined in refinement tasks. Additionally, future work could address the limitations we mentioned by involving a larger sample size to capture broader preferences, conducting system-based tests with generated outputs for a more comprehensive experience (e.g., actual system implementation or Wizard-of-Oz studies), and exploring iterative refinements in real-world contexts to identify overlooked task types.

6 Conclusion

This paper explored preferences for annotations, scribbles, and text prompts in GenAI image tools for refinement tasks, addressing gaps in text-centric inputs and the predominant focus of prior research on the ideation phase rather than refinement. Our preliminary digital paper-based study with seven professional designers across five refinement task types revealed that input preferences varied based on the task and user goals. The findings indicate that annotations were preferred for making spatial adjustments and referencing elements within images. Scribbles, often used alongside other methods, were effective in specifying attributes such as shape, size, and position. Text prompts were favored for tasks requiring detailed instructions or allowing greater creative input from GenAI. However, each method had its drawbacks: annotations and scribbles raised concerns about AI misinterpretation, while crafting effective text prompts demanded significant effort. Our insights highlight the need for interfaces that balance control and flexibility, considering specific design tasks and diverse work setups while moving beyond traditional text-prompting methods. As AI systems improve in interpreting scribbled or handwritten annotated visual inputs, HCI research could explore effective ways to enhance user intent communication within interfaces. Future research could investigate whether to prioritize the development of text-based prompt creation tools or focus on alternative input methods that better support visually oriented designers and users at different stages of the design process and creative workflow.

References

[1] DALL·E 3. 2025. DALL·E 3. https://openai.com/index/dall-e-3/
[2] Samsung Galaxy AI. 2024. Use AI editing tools in Gallery on your Galaxy phone or tablet. https://www.samsung.com/us/support/answer/ANS10000934/
[3] 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

[4] Liuqing Chen, Qianzhi Jing, Yixin Tsang, Qianyi Wang, Ruocong Liu, Duowei Xia, Yunzhan Zhou, and Lingyun Sun. 2024. AutoSpark: Supporting Automobile Appearance Design Ideation with Kansei Engineering and Generative AI. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, Pittsburgh PA USA, 1–19. https://doi.org/10.1145/3654777.3676337

[5] DaEun Choi, Sumin Hong, Jeongeon Park, John Joon Young Chung, and Juho Kim. 2024. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–25. https://doi.org/10.1145/3613904.3642794

[6] 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*. ACM, San Francisco CA USA, 1–17. https://doi.org/10.1145/3586183.3606777

[7] Hai Dang, Frederik Brudy, George Fitzmaurice, and Fraser Anderson. 2023. WorldSmith: Iterative and Expressive Prompting for World Building with a Generative AI. 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/10.1145/3586183.3606772

[8] Dreamstudio. 2024. DreamStudio. https://dreamstudio.ai/

[9] K. J. Kevin Feng, Maxwell James Coppock, and David W. McDonald. 2023. How Do UX Practitioners Communicate AI as a Design Material? Artifacts, Conceptions, and Propositions. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 2263–2280. https://doi.org/10.1145/3563657.3596101

[10] Firefly. 2024. Adobe Firefly - Free Generative AI for creatives. https://www.adobe.com/products/firefly.html

[11] 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*. ACM, Hamburg Germany, 1–20. https://doi.org/10.1145/3544548.3580999

[12] DALL·E Inpainting. 2024. Editing your images with DALL·E | OpenAI Help Center. https://help.openai.com/en/articles/9055440-editing-your-images-with-dall-e

[13] Midjourney Inpainting. 2024. Vary Region + Remix. https://docs.midjourney.com/docs/vary-region

[14] Qianzhi Jing, Tingting Zhou, Yixin Tsang, Liuqing Chen, Lingyun Sun, Yankun Zhen, and Yichun Du. 2023. Layout Generation for Various Scenarios in Mobile Shopping Applications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3544548.3581446

[15] Colin Lea, Zifang Huang, Jaya Narain, Lauren Tooley, Dianna Yee, Dung Tien Tran, Panayiotis Georgiou, Jeffrey P Bigham, and Leah Findlater. 2023. From User Perceptions to Technical Improvement: Enabling People Who Stutter to Better Use Speech Recognition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. https://doi.org/10.1145/3544548.3581224

[16] Seung Won Lee, Tae Hee Jo, Semin Jin, Jiin Choi, Kyungwon Yun, Sergio Bromberg, Seonghoon Ban, and Kyung Hoon Hyun. 2024. The Impact of Sketch-guided vs. Prompt-guided 3D Generative AIs on the Design Exploration Process. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. https://doi.org/10.1145/3613904.3642218

[17] Zhaohui Liang, Xiaoyu Zhang, Kevin Ma, Zhao Liu, Xipei Ren, Kosa Goucher-Lambert, and Can Liu. 2024. StoryDiffusion: How to Support UX Storyboarding With Generative-AI. https://doi.org/10.48550/arXiv.2407.07672 arXiv:2407.07672.

[18] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3526113.3545621

[19] 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*. ACM, Pittsburgh PA USA, 1955–1977. https://doi.org/10.1145/3563657.3596098

[20] Midjourney. 2024. Midjourney. https://www.midjourney.com/home?callbackUrl=%2Fexplore

[21] Hyerim Park and Malin Eiband. 2024. Designing for Visual Thinkers: Overcoming Text-Centric Limitations in GenAI Tools. https://doi.org/10.5281/zenodo.14186390

[22] Hyerim Park, Joscha Eirich, Andre Luckow, and Michael Sedlmair. 2024. "We Are Visual Thinkers, Not Verbal Thinkers!": A Thematic Analysis of How Professional Designers Use Generative AI Image Generation Tools. In *Nordic Conference on Human-Computer Interaction*. ACM, Uppsala Sweden, 1–14. https://doi.org/10.1145/3679318.3685370

[23] Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2024. DesignPrompt: Using Multimodal Interaction for Design Exploration with Generative AI. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 804–818. https://doi.org/10.1145/3643834.3661588

[24] Xinyu Shi, Ziqi Zhou, Jing Wen Zhang, Ali Neshati, Anjul Kumar Tyagi, Ryan Rossi, Shunan Guo, Fan Du, and Jian Zhao. 2023. De-Stijl: Facilitating Graphics Design with Interactive 2D Color Palette Recommendation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3544548.3581070

[25] Kihoon Son, DaEun Choi, Tae Soo Kim, Young-Ho Kim, and Juho Kim. 2024. GenQuery: Supporting Expressive Visual Search with Generative Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3613904.3642847

[26] Macy Takaffoli, Sijia Li, and Ville Mäkelä. 2024. Generative AI in User Experience Design and Research: How Do UX Practitioners, Teams, and Companies Use GenAI in Industry?. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 1579–1593. https://doi.org/10.1145/3643834.3660720

[27] 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

[28] Michihiko Ueno and Shin'ichi Satoh. 2021. Continuous and Gradual Style Changes of Graphic Designs with Generative Model. In *Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21)*. Association for Computing Machinery, New York, NY, USA, 280–289. https://doi.org/10.1145/3397481.3450666

[29] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–21. https://doi.org/10.1145/3613904.3642803

[30] Jeremy Warner, Kyu Won Kim, and Bjoern Hartmann. 2023. Interactive Flexible Style Transfer for Vector Graphics. 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/10.1145/3586183.3606751

[31] Di Wu, Zhiwang Yu, Nan Ma, Jianan Jiang, Yuetian Wang, Guixiang Zhou, Hanhui Deng, and Yi Li. 2023. StyleMe: Towards Intelligent Fashion Generation with Designer Style. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3544548.3581377

[32] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. 2018. Generative Image Inpainting with Contextual Attention. https://doi.org/10.48550/arXiv.1801.07892 arXiv:1801.07892 [cs].

[33] Hongbo Zhang, Pei Chen, Xuelong Xie, Chaoyi Lin, Lianyan Liu, Zhuoshu Li, Weitao You, and Lingyun Sun. 2024. ProtoDreamer: A Mixed-prototype Tool Combining Physical Model and Generative AI to Support Conceptual Design. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3654777.3676399

A User Input Examples: Guiding GenAI for Refinements

This appendix provides examples from the user study, illustrating how participants used their preferred input methods—text prompts, annotations, and scribbles—to guide GenAI image tools in various refinement tasks. These methods were not strictly separate; at times, participants considered combining multiple inputs within a single task, thinking that this approach might lead to more effective outcomes.
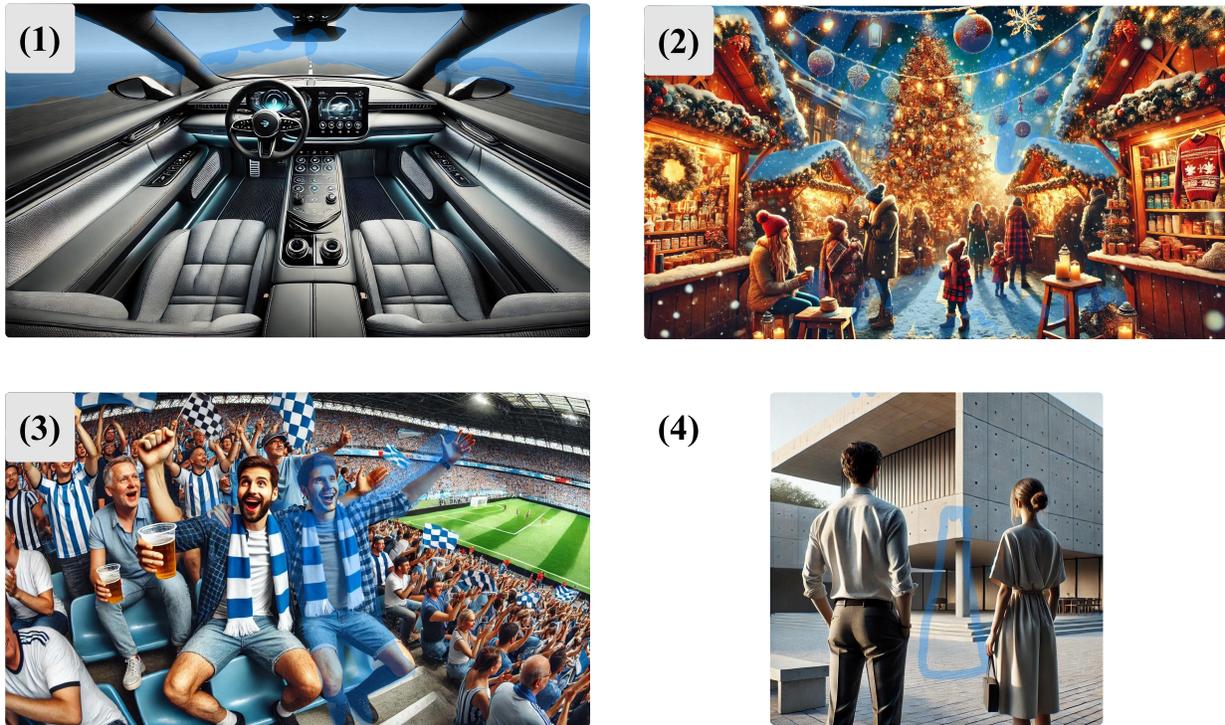
A.1 Text Prompts

Figure 2: Text Prompt Input Examples: (1) and (2) involve inpainting (highlighted in blue) applied to large areas for background changes, such as adding a forest (*"Add a forest as a background outside the car"*) or increasing snowfall (*"Make the snowfall bigger and heavier"*). (3) A participant used inpainting to modify a man into a girlfriend figure with the text prompt, *"Change this man into a woman with a girlfriend look."* The girlfriend figure allowed room for AI's creative interpretation, as the prompt required less precise instructions. (4) A participant marked an area and provided the text prompt, *"Add a Christmas tree decorated with ornaments and lighting. The decoration should be modern and chic, with yellow and white lights. It should match the building behind."* The instructions were relatively lengthy and detailed. Some participants relied solely on text prompts without inpainting to adjust the overall mood of the image, such as removing the yellow tone entirely.
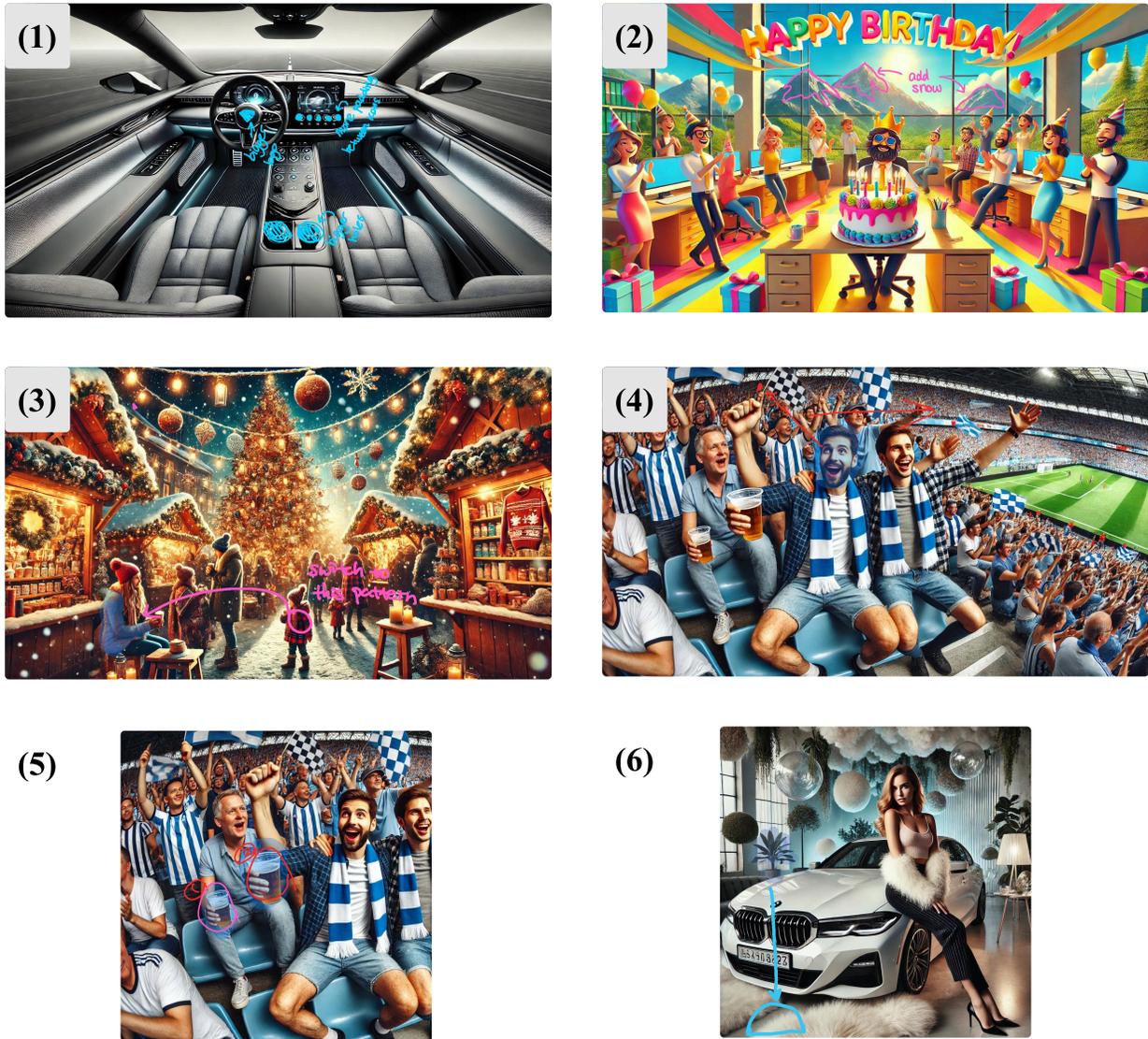
A.2   Annotations

Figure 3: Annotation Input Examples: Annotations involve text or visual symbols (e.g., circles, arrows, lines, numbering), with arrows and circles being the most commonly used. At times, annotations were combined with scribbles and text prompts. (1) A participant used scribbles to indicate the size of a logo on the steering wheel or buttons on the display, paired with text annotations like *"bigger"* or *"more padding."* (2) Scribbles indicated snow distribution on mountains, clarified by a text annotation reading *"Add snow."* (3) An arrow and text annotation (*"Switch to this pattern"*) were used to apply a checkered pattern from one person's clothing to another's in the image. (4) A participant used an arrow and dot to indicate a change in eye direction, with a text prompt stating *"Change the eyes to look in the direction of the dots."* (5) Numbering (e.g., *"1"* and *"2"*) labeled two beer glasses, with a text prompt stating, *"Make 1 and 2 the same size."* (6) Visual symbols, including an arrow and a circle, showed how to move a plant to a new position in the image.
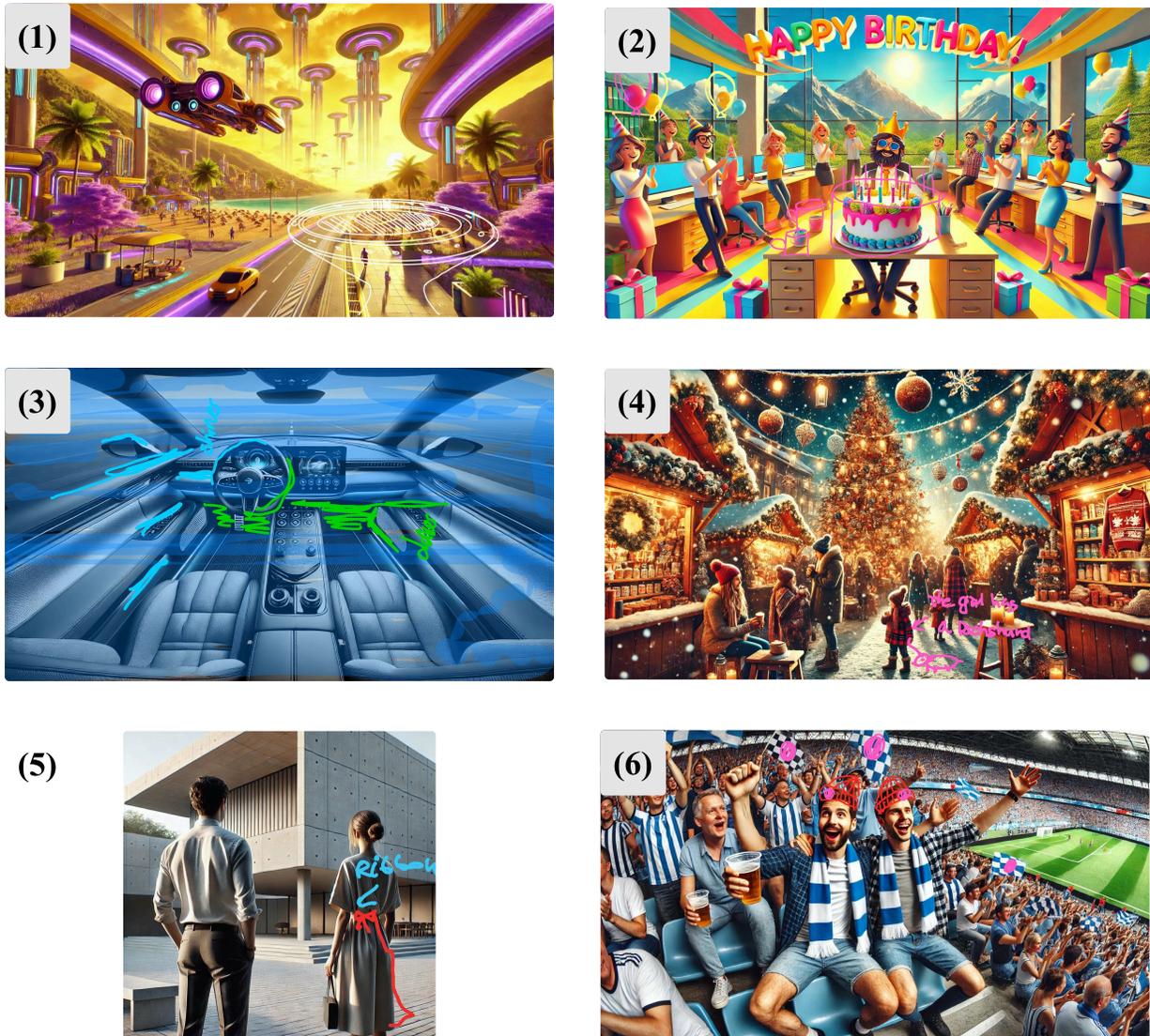
A.3 Scribbles

Figure 4: Scribble Input Examples: Most users combined scribbles with text prompts or annotations to add details. (1) A scribble depicted the shape, size, and rough design of a flying car station to be added. (2) A participant used scribbles to specify the intended size of a cake, a balloon, and a new element (a cup) on a desk without additional input. (3) Scribbles highlighted specific regions of a car for lighting effect adjustments, paired with annotations specifying areas to brighten (*in blue*) and darken (*in green*). (4) A participant used rough scribbles to indicate the size and orientation of a dachshund, complemented by an annotation labeling it as "*dachshund*." (5) Scribbles showed the desired position of a ribbon and modifications to a skirt shape, with an annotation specifying "*ribbon*." (6) A scribble outlined a beanie design without a pompom, while annotations and text prompts detailed adding a "*Tottenham logo*" to the pink-circled area.