透過可追溯的提示詞提升環境設計中圖像精修的可控性

王文凡*
vann@cmlab.csie.ntu.edu.tw
國立台灣大學
台北, 台灣

李婷穎*
tylee@cmlab.csie.ntu.edu.tw
國立台灣大學
台北, 台灣

呂建廷
B09902109@csie.ntu.edu.tw
國立台灣大學
台北, 台灣

徐晢偉
yaweihsu1234@gmail.com
國立台灣大學
台北, 台灣

彭煦楠
R12944063@ntu.edu.tw
國立台灣大學
台北, 台灣

陳譽
r11922026@ntu.edu.tw
國立台灣大學
台北, 台灣

陳彥仰
mikechen@csie.ntu.edu.tw
國立台灣大學
台北, 台灣

陳炳宇
robin@ntu.edu.tw
國立台灣大學
台北, 台灣

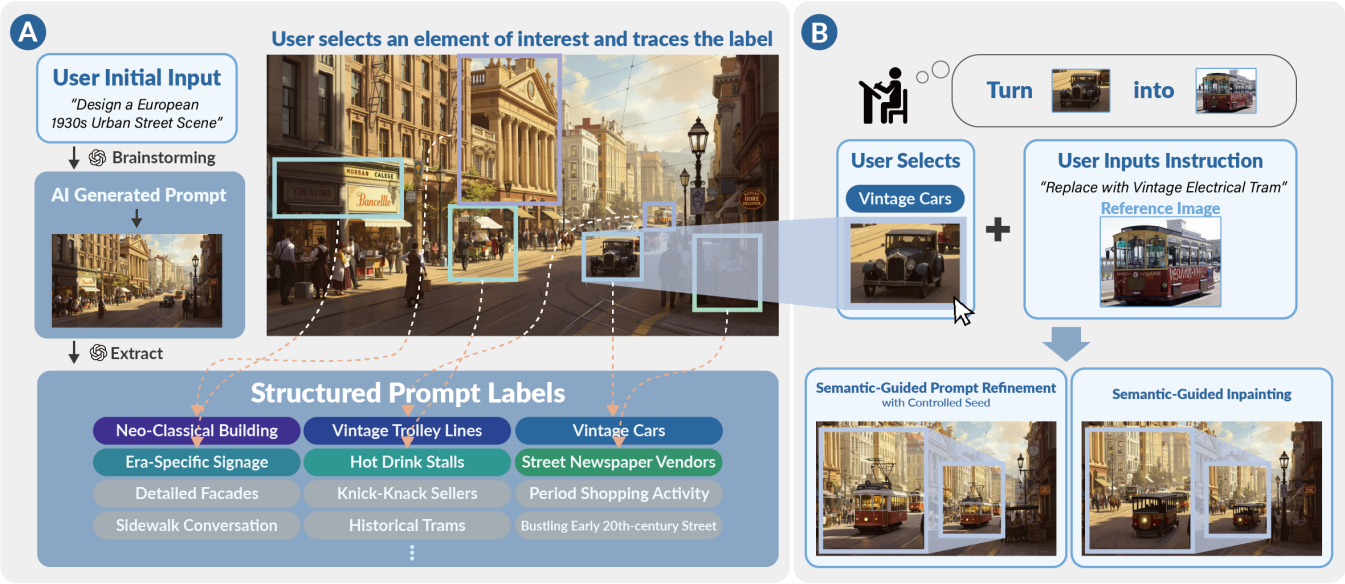

Figure 1: GenTune, a human-centered generative AI system with traceable prompts for controllable image refinement in environment design. (A) Begins with a user's initial input-*Design a European 1930s Urban Street Scene*-which is expanded by a Brainstorming LLM into a structured prompt for image generation. A label extraction LLM then extracts key elements to establish prompt – image element correspondences, supporting user understanding. (B) During refinement, the user selects a region of interest to reveal its associated label-*Vintage Cars*. Upon entering a refinement instruction-*Replace with Vintage Electrical Tram*—along with a reference image, GenTune applies both semantic-guided prompt refinement with controlled seed and semantic-guided inpainting to generate updated results.

Abstract

在娛樂產業中，場景設計師為遊戲、電影與電視製作 2D 與 3D 場景，他們不僅需要對細節有精細掌控，也需要維持整體畫面的一致性。隨著科技發展，設計師們越來越需要在工作流程中使用生成式人工智慧（Generative AI），例如，使用大型語言模型（LLMs）來增強文字生成圖像的提示詞

*Both authors contributed equally as first author.

CGW '25, Taipei, Taiwan,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

（prompts），再反覆迭代修改提示詞與局部修補（inpainting）來精修圖像。然而，我們針對 10 位設計師進行的初步研究顯示了兩項主要挑戰：（1）LLM 生成的提示語非常冗長複雜，致使設計師難以理解並且找出對應特定視覺元素的關鍵詞；（2）雖然局部修補能編輯特定區域，但在維持圖像整體一致性與合理性方面是一大挑戰。

基於這些觀察，我們提出 GenTune，一種強化人與 AI 協作的系統，透過清楚地呈現 AI 生成提示詞與圖像內容之間的對應關係，幫助設計師更有效地進行編輯。GenTune 系統讓設計師可以選取生成圖像中的任意物件，追溯其對應的提示詞標籤，並透過這些標籤引導圖像進行精確且整體一致的優化。在一項針對 20 位設計師的總結性研究中，GenTune 在提示詞與圖像的理解度、編輯品質與效率、以及整體滿意度方面，均較現行方法有顯著提升（皆達顯著水準，$p < .01$）。

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; **User centered design**.

Keywords

Generative AI, Human-Centered AI, Environment Design, Creativity Support Tool, Visual Exploration, Traceable Prompt

ACM Reference Format:
王文凡, 李婷穎, 呂建廷, 徐晢偉, 彭煦楠, 陳譽, 陳彥仰, and 陳炳宇. 2025. 透過可追溯的提示詞提升環境設計中圖像精修的可控性. In *Proceedings of July 10–11, 2025 (CGW '25).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Environment designers in the entertainment industry craft the visual and spatial worlds that audiences experience in games, animations, films, and TV shows [2, 26, 32, 45]. Their work typically occurs during pre-production, where they collaborate with art directors to develop 2D and 3D concepts that define a project ' s visual direction [45]. These designs often serve as blueprints for modeling and VFX teams, or in smaller or stylized productions, are used directly in final scenes [12, 18]. The traditional workflow involves two phases: early ideation—researching, brainstorming, and drafting variations—then final refinement, where approved concepts are polished into production-ready assets [1, 26, 32].

With the rise of generative AI (GenAI) tools, environment designers increasingly integrate them into their workflows [28, 29, 44, 53]. They collaborate with multimodal large language models (MLLMs) to craft and refine prompts, which serve as input for text-to-image (T2I) models. The recent release of ChatGPT-4o's image generation[1] further signals the growing adoption of AI for advanced image editing. GenAI is now used across ideation, inspiration, client communication, and even production-level outputs [19, 37, 53]. This shift has raised industry expectations—designers are expected to iterate faster and deliver higher-quality visuals. In response, recent research has explored deeper GenAI integration, such as prompt-tuning [8, 54, 55] and multimodal models for rapid ideation [10, 35, 53].

However, as designers move to refinement, they often need to revisit, modify, or build upon initial AI-generated outputs—both prompts and images. Current GenAI tools treat prompts as opaque, one-shot inputs, making it hard to trace which parts correspond to specific visual elements. While MLLMs like ChatGPT show promise in image editing and spatial coherence, they struggle with the complexity of environment design, and their automated nature limits designer control and restricts expressive intent. As a result, designers struggle with control and precision in the refinement process, often resorting to time-consuming trial-and-error cycles [6, 9, 51, 53]. In this work, we address these challenges with a human-centered approach tailored to the specific needs of environment designers, supporting better understanding and refinement of generated outputs during the pre-production process.

To better understand the challenges environment designers face with GenAI tools, we conducted a formative study with 5 professionals and 5 design students who regularly use them in their workflows. Through workflow analysis and in-depth interviews, we identified two core challenges: a lack of understanding of how text prompts relate to generated image elements, and limited control during the refinement process. Environment design involves complex spatial and visual composition across both macro (layout) and micro (detail) scales. While designers often rely on LLM to craft long, detailed prompts to guide generation, they frequently struggle to trace how specific parts of the prompt influence the output—making refinement a frustrating trial-and-error process. For local refinement, approaches like inpainting [59] often lead to inconsistencies in lighting, style, or context. More technical solutions, such as ComfyUI [16] with ControlNet [60], offer finer control but are too complex and misaligned with designers' workflows.

To address the specific needs of environment designers, we present GenTune, a human-centered GenAI system designed to enhance prompt-to-image interpretability and the refinement process. For the initial image generation, GenTune builds on prior work [10, 53] by incorporating a brainstorming module that transforms simple user inputs into detailed prompts and generates four diverse visual outputs. GenTune introduces two key modules following the initial generation:

(1) **Traceable Prompt:** Designers can select an area of interest in the image to reveal a corresponding label extracted from the structured prompt used to generate the image (Fig. 1-A). The full prompt segment associated with the label is also displayed, helping designers understand how specific parts of the prompt relate to visual elements.
(2) **Semantic-Guided Refinement:** GenTune allows designers to precisely refine the image based on a selected area of interest (Fig. 1-B). The system supports three refinement modes: refining the image element associated with the selected label, modifying only the selected region, or comparing both and choosing the preferred result. Designers can input refinements via natural language and reference images. Additionally, the system suggests refinement options based on both the selected element and the overall image context. These features enable designers to achieve their desired results efficiently.

We conducted a summative study with 15 professional environment designers and 5 design students, all experienced with GenAI tools, to evaluate the effectiveness of GenTune. The study included

[1] ChatGPT, https://openai.com/index/introducing-4o-image-generation/

a within-subjects experiment simulating real-world generative image refinement tasks, comparing GenTune to a baseline system without its two key modules. Results showed that, compared to the baseline, participants significantly better understood the prompt–image connection ($p = 0.003$), found GenTune more effective for refinement ($p = 0.002$), produced higher-quality outputs ($p = 0.003$), and reported greater satisfaction ($p < 0.001$). An open-ended task followed, in which participants applied GenTune to their own past or ongoing GenAI projects and compared it with their typical workflow. Participants significantly preferred GenTune in terms of refinement efficiency, quality, and creativity support ($p < 0.001$ for all), and also reported feeling more in control and more satisfied overall ($p < 0.001$).

In summary, the major contributions of this work are:

- A formative study investigating the end-to-end GenAI workflow of professional environment designers, identifying key challenges and specific needs in refining generated images.
- The design and implementation of GenTune, a human-centered GenAI system that supports targeted, semantic image refinements via natural language or image references. GenTune features Traceable Prompt and Semantic-Guided Refinement to help designers better understand and control AI outputs.
- A comprehensive multi-stage evaluation showing that GenTune significantly improves designers' ability to interpret, control, and refine generative outputs. This includes: (1) a within-subjects experiment, and (2) an open-ended task using with production projects.

2 FORMATIVE STUDY

We conducted a formative study to investigate how environment designers currently use GenAI design tools, explore their workflows, and identify the challenges they face.

2.1 Participants

Our formative study included 5 professional environment designers from the game and animation industries (3–8 YoE, mean = 4.8), and 5 design students with at least six months of intensive GenAI experience. Participants were compensated 15 $USD.

2.2 Study Procedure

The study consisted of three parts. We began with a 20-minute in-depth interview covering: (1) participants' goals when using GenAI tools, (2) their strategies and workflows, and (3) past experiences using GenAI tools in projects, including challenges they encountered. A 30-minute workflow observation session followed to gain deeper insight into their design strategies. Participants used their preferred GenAI tools to design a Medieval Chinese Science Center. Finally, a 20-minute post-task interview focusing on their refinement strategies and difficulties.

2.3 Findings

2.3.1 Roles of GenAI in Environment Design. Designers commonly use GenAI tools for early-stage ideation, particularly when working with unfamiliar design specifications. As one participant noted, " *It's common that we cannot find specific design references to the topic, but image generation tools can provide that* " (P4). GenAI also serves as a source of visual reference, and designers draw inspiration from mood, atmosphere, spatial layout, and textures. " *Sometimes I just want to see how AI would handle certain architectural structures* " (P1). As GenAI tools become more widespread, art directors and clients increasingly expect environment designers to use them for faster visual communication and decision making. This shift has raised expectations for speed and quality. As one participant described, " *Clients think GenAI is powerful and expect high-quality revisions daily, not every few days like before* " (P3). Despite this pressure, GenAI output still requires significant manual refinement. As one designer explained, " *AI rarely captures exactly what the client wants—we still do a lot of post-processing to meet their expectations* " (P4).

2.3.2 Current GenAI Refinement Workflow and Key Challenges. Despite our participants used various image generation tools, their workflows were similar: designers often crafted long and intricate prompts to specify the desired elements, many used LLMs to generate initial prompts, then refined the images iteratively from global structures to local details.

A key challenge was the lack of clear mapping between prompt text and visual elements, making it hard to determine which parts of the prompt control specific image features. " *I often have to search for a long time just to find the section I need to edit* " (P10). Even then, designers were unsure if their edits had the intended effect. " *Many times the new image makes me question whether I actually edited the right part* " (P2). As a result, designers often rely on trial and error. Even with LLM support, ambiguity persists. " *I feel like I've described exactly how to change it, but the AI still doesn't understand* " (P9). Moreover, generative images also include unexpected elements not mentioned in the prompt. These challenges underscore the inefficiencies of prompt-based refinement, especially in complex environment design workflows.

Another key issue is structural consistency, which is critical in environment design, especially when presenting to stakeholders. However, general-purpose GenAI tools often fail to preserve structural coherence. " *If we need to revise an AI image for clients, we usually just edit it manually in Photoshop—it's faster* " (P4). Some designers tried technical workflows like using structural conditioning in ComfyUI [16], but found them ineffective for complex scenes. " *Even if the structure stays the same, all the textures go off* " (P1). Others found these tools too complex to use. As P3 put it, " *I'm an artist, not an engineer—why should I have to use this (ComfyUI)?* "

To modify local details without affecting other regions, some designers used prompt-based inpainting [59], a common feature in GenAI tools. This involves selecting a region and entering a replacement prompt. However, results were often inconsistent—mismatched lighting, style, proportion, or contextual coherence, such as historical or spatial consistency. " *Inpainting often gives me strange results—I've tried many different prompts, but still didn't get what I wanted* " (P6). Others opted for manual editing, but the lack of editable layers made this difficult. " *It's faster to repaint the area from scratch* " (P3).

These challenges hinder environment designers from effectively refining GenAI images—despite growing expectations for seamless integration in industry workflows.

2.4 Design Goals

Base on the finding, we proposed three design goals for our system:

- **DG1: Transparent Prompt-to-Image Mapping.** The system should enable clear, interactive mappings between prompt text and corresponding visual elements, helping environment designers understand how specific prompts influence outputs. This facilitates more precise and intentional control during the refinement process.
- **DG2: Coherent Refinement from Global to Local.** The system should align with designers' workflows by supporting iterative refinement from global structure to local details, preserving visual coherence while enabling targeted, flexible edits.
- **DG3: Intuitive and Predictable Control Workflow.** The system should provide intuitive interaction flow that promote a strong sense of control and predictability, making refinement outcomes more understandable and reliable. This reduces trial-and-error and enhances designers' creative confidence.

3 SYSTEM & IMPLEMENTATION

We present GenTune, a human-centered generative AI system that enhances interpretability and user control in image refinement for environment design. GenTune helps designers quickly identify areas of interest and supports precise, progressive refinement.

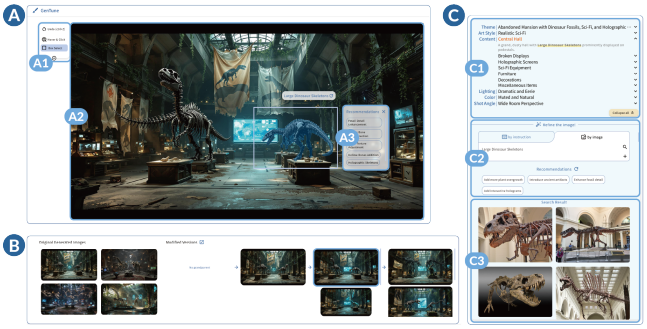

Figure 2: GenTune includes: (A) a main image panel with a region selection mode selector (A1), selected region display (A2), and corresponding labels with refinement suggestions (A3); (B) a version history view showing the relationship between initial and refined images; and (C) a refinement panel with a structured prompt view (C1), a text input dialog and suggestions (C2), and reference image search (C3).

3.1 System Overview

GenTune's image generation system draws inspiration from conversational generation systems [23, 53, 56], which support iterative, dialogue-driven workflows. It features a brainstorming module that transforms simple user input into high-quality prompts and generates four initial images through a conversational interface. Once an image is selected, it enters GenTune's main interface (Fig. 2), which features two core modules designed to support our key goals.

3.1.1 Traceable prompt. GenTune structures prompts into six key categories—theme, art style, content, lighting, color, and shot angle—reflecting the most emphasized aspects of environment design from our formative study.

To help designers identify which parts of the prompt correspond to visual elements, GenTune supports two selection modes (Fig. 2-A1): (1) hover-and-click, which auto-detects elements and overlays a blue segmentation mask, and (2) box selection, which matches the best segmented region within the selected area (Fig. 2-A2). Once an element is selected, a label appears above it, indicating the corresponding prompt keyword, and the relevant section in the prompt overview panel automatically expands (Fig. 2-C1). Labels are organized in a tree structure, allowing parent subcategories to expand when a label is selected. These labels are derived from the content category, which most directly maps to identifiable visual elements.

3.1.2 Semantic-guided refinement. After tracing a label, designers can refine the image using text, image input, or both. They may enter natural language instructions in the dialogue box and provide a reference image via the embedded search engine or by uploading directly through the right-hand panel (Fig. 2-C2).

GenTune supports three refinement options, each operating at a different scope—from global to local:

- **Global refinement (no selection required)**. This mode allows designers to edit the entire image without selecting a specific label, using only a text instruction or reference image. It builds on prior conversational image editing systems and supports broad visual changes.
- **Semantic-guided prompt refinement with controlled seed (requires selection)**. In this novel method, GenTune makes targeted edits to the original prompt based on the designer's input and the selected label, then regenerates the image using the same seed. This allows changes to apply only to the intended element while preserving overall coherence [21, 38]. For example, the designer selects the element labeled " Vintage Cars " and provides a reference image of a vintage tram. GenTune replaces cars with trams and adds overhead wires to ensure contextual consistency, while keeping the rest of the image largely unchanged (Fig. 1-B, Fig. 6-B). This design choice helps align AI-driven refinement with designer intent. Although minor regions of the image may shift, environment designers typically prioritize conceptual consistency and visual coherence over exact pixel-level accuracy—a preference that professionals in our user evaluation later validated.
- **Semantic-guided inpainting (requires selection)** This option enables more localized edits. As noted in our formative study, traditional inpainting methods require precise input and often yield inconsistent results. In contrast, GenTune accepts simple natural language commands (e.g., " add some merchants ") and generates context-aware prompts using the selected label and original prompt. This allows the inserted content to remain stylistically and semantically consistent with the overall scene.

For each refinement, GenTune generates four image variations. It offers three modes: seed mode and inpainting mode, each producing four results, and mixed mode, which returns two from each.

This allows designers to compare outcomes across strategies—especially useful when they are unsure about the scale of change or want to explore stylistic trade-offs.

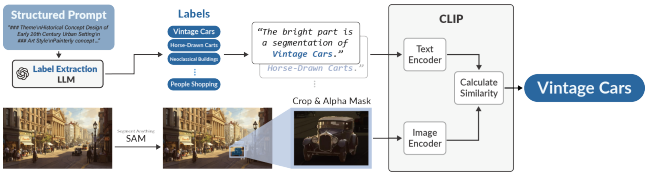

Figure 3: Pipeline for prompt-element correspondence. Label Extraction LLM extracts labels from the structured prompt and used as text inputs to CLIP. When the user selects a region, it is segmented by SAM, cropped with a bounding box, alpha-masked, and fed into CLIP. The CLIP model then calculates text-image similarity to identify the label most closely associated with the selected region.

3.2 User Interface

GenTune features a web-based interface with three main pages: (1) a front page for initial input, (2) an overview page displaying four generated images, and (3) the main interface with GenTune's two core modules (Fig. 2). In the main interface, users can hover or draw a box to reveal element labels on the image (Fig. 2-A2), click the refresh icon to cycle through alternatives, and use the checkmark icon to view label-based suggestions. The right panel contains the prompt overview (Fig. 2-C1) and a dialog box (Fig. 2-C2) for entering text or uploading reference images. Users can switch between Mixed, Seed, and Inpainting modes via the mode button. Prompt suggestions appear below and can be refreshed based on the user's input. During refinement, a "Generating image" message appears in the bottom-right corner, which can be clicked to view progress or wait for a result pop-up. The bottom of the interface displays an image iteration tree (Fig. 2-B), showing thumbnails of the hierarchical relationships of each image, to help track and revisit edits.

4 SUMMATIVE STUDY

Our summative study evaluates how GenTune's two core modules support professionals in understanding, controlling, and refining generative AI outputs within their workflows. We conducted two complementary experiments:

(1) **A within-subjects experiment simulating real-world generative image refinement workflows.** To benchmark GenTune, we developed a baseline system with a similar UI that reflects common industry workflows. It included: (1) Conversational Image Editing, where designers edited scenes using natural language or reference images via an LLM; and (2) Basic Inpainting, where designers manually selected regions, entered prompts, and applied localized edits using Flux Fill. The baseline excluded GenTune's two core modules, requiring participants to manually read and iteratively adjust prompts and use inpainting tools for fine-grained control. In both conditions, the full structured prompt was displayed (Fig. 2-C1).

(2) **An open-ended task.** Designers applied GenTune to their own projects involving generative AI tools and compared the experience to their usual workflows.

We focus on three key aspects:

- A1: Prompt–image interpretability
- A2: Refinement effectiveness and output quality
- A3: Alignment with expectations and control

4.1 Study Design

4.1.1 Task overview. In the first task, participants completed an image refinement exercise using both the baseline system and GenTune. Each involved one of two assigned design topics with a pre-generated image: (1) The Hanging Gardens of Neo-Babylon or (2) The Floating Monastery of the Himalayas. Participants completed four rounds of refinement—one global and three local—based on client-style instructions. Local edits specified regions to modify, and participants chose the refinement order freely. Before each edit, participants identified the prompt corresponding to the element they wished to modify, then selected one of four generated image candidates to continue refining, with up to two iterations per edit. Selection was based on consistency (style, lighting, color, structure, context), aesthetics, and alignment with client intent. Final images were served as references for client communication and future development. Design topics and refinement instructions were validated by two professional art directors. Figure **??** shows an example workflow for the first topic under both conditions.

For the second open-ended task, participants began by describing the workflows of recent environment design projects they had worked on. They then recreated two to three of these projects using GenTune, iteratively refining each image until satisfied, and selecting a final result that aligned with their creative intent and was suitable for client communication.

4.1.2 Measurements. The post-condition questionnaire for the first task evaluated three key aspects: image–prompt understanding, refinement effectiveness and quality, and overall user experience. It also included a NASA-TLX to assess perceived workload. All questions and results are shown in Figure 4. Responses were collected using a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). We adopted a self-report approach, consistent with prior HCI and creativity research [30, 34, 43, 49], and analyzed the data using the Wilcoxon signed-rank test [57], while NASA-TLX scores were analyzed using paired t-tests.

For the second task, the questionnaire asked participants to rate their preference between GenTune and their previous approach across three core aspects. Responses were recorded on a 7-point Likert scale (1 = strong preference for their original workflow, 7 = strong preference for GenTune). Questions and results are shown in Figure 5. We used a one-sample Wilcoxon signed-rank test to assess whether responses significantly differed from the neutral midpoint (4), appropriate for ordinal data from Likert-scale preference questionnaires [11, 41, 50].

In-depth interviews complemented the second open-ended task by offering qualitative insights into how professionals engaged

with GenTune in real-world workflows. We focused on how GenTune influenced the understanding of the prompt-image relationship, and how GenTune differed from their previous workflows, editing methods, refinement efficiency and quality.

4.1.3 Participants. We recruited 15 professional environment designers with 1 to 21 YoE (Mean = 6.60) from various industries, including game (P1, P4, P6, P11-P13, P17), animation (P3, P5, P14, P21), film (P15-P16), industrial design (P20) and freelancing (P18-19). We also included 5 design students (P2, P7 – P10) who had at least six months of intensive experience with GenAI design tools on environment design projects.

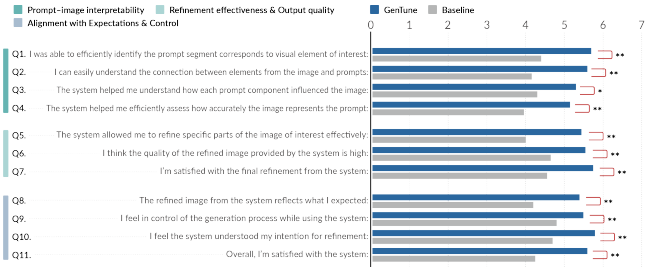

Figure 4: Survey results from the within-subject task. *: p < .05 and **: p < .01.

5 RESULTS & FINDINGS

In this section, We report findings organized by our three evaluation aspects (A1 – A3), combining results from both the controlled experiment and open-ended task.

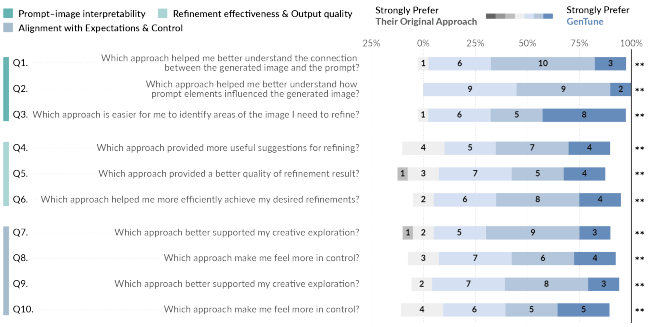

Figure 5: User preference distribution for their original approach vs. GenTune . **: p < .01.

5.1 A1: Prompt–Image Interpretability

Participants using GenTune showed significantly improved prompt-to-image interpretability. In the within-subjects study, participants found it significantly easier to identify which prompts corresponded to specific visual elements (Fig. 4, Q1; Mean_diff = 1.35, p = 0.002), and to assess how accurately the generated image reflected the prompt (Fig. 4, Q4; Mean_diff = 1.2, p < 0.001). In the open-ended task, 95% preferred GenTune for identifying which parts of the image to refine (Fig. 5, Q3; Mean = 6.00, p < 0.001). Many participants noted that GenTune saved them from reading detailed prompts to find what they wanted to modify (P2-8, P11-17, P19-21). For example, " *The highlights clearly show what needs to be changed—I don't have to dig through the prompt* " (P3).

Participants found it significantly easier to understand the connection between prompts and visual outputs using GenTune. In the within-subjects study, they more effectively connected image elements to their corresponding prompts (Fig.4, Q2, Mean_diff = 1.1, p = 0.003), and better understood how each prompt influenced the image (Fig.4, Q3, Mean_diff = 0.7, p = 0.03). In the open-ended task, 95% and 100% of participants, respectively, preferred GenTune for these aspects (Fig. 5, Q1, Mean = 5.75, p < 0.001; Q2, Mean = 5.65, p < 0.001). Many found that selecting a label clarified what would be affected (P2-6, P8-13, P16-P21). " *The labels are intuitive—once it's highlighted, I know what the system will affect. I don't have to second-guess or worry about unintended changes* " (P6).

However, interpretability decreased in scenes with many similar elements. For example, P11, designing a futuristic plant lab filled with bottles and specimens, found it hard to distinguish between overlapping labels: " *Each prompt candidate seemed plausible, but I couldn't tell which one was actually correct.* "

Overall, these findings highlight GenTune's strength in bridging the gap between textual prompts and visual outputs, enabling designers to interpret AI-generated images with greater clarity.

5.2 A2: Refinement effectiveness and output quality

5.2.1 Refinement efficiency and precision. In the within-subjects study, participants rated GenTune significantly more effective for image refinement (Fig.4, Q5; Mean_diff = 1.2, p = 0.002), and preferred it for achieving desired result more efficiently in the open-ended study (Fig.5, Q6; Mean = 5.7, p < 0.001), with 90% favoring GenTune. Most reported that semantic-guided refinement accurately targeted the areas they wanted to change (P2 – 4, P6 – 10, P12 – 21) and reduced trial and error. As P12 explained, " *Before, I kept revising prompts because I wasn't sure what they referred to. With labeled highlights, I know exactly what each part means—no more guesswork.* " Some participants found GenTune especially efficient for editing multiple similar elements (P6, P9, P13–14, P17), highlighted its efficiency in adding or replacing elements (P2–3, P6, P10, P12–13, P19). As P9 noted, " *Modifying multiple elements was much faster—GenTune could update all label-related parts at once, previously a slow, manual task in Photoshop.* "

5.2.2 Refinement quality. Images refined with GenTune were rated significantly higher in quality than the baseline (Fig.4, Q6; Mean_diff = 0.9, p = 0.003), with 80% preferring its quality in the open-ended task (Fig.5, Q5; Mean = 5.4, p < 0.001). Participants also reported greater satisfaction with the final results than the baseline(Fig.4, Q7; Mean_diff = 1.45, p < 0.001). Many were impressed because they typically relied on inpainting or manual editing (P3, P5 – 8, P11–14, P16–17), while GenTune's semantic-guided prompt refinement achieved better results with minimal structural disruption. As P20 said, " *GenTune let me control the structure while making precise edits—other tools made unpredictable changes.* " One exception was P14, who preferred the aesthetic quality of their previous workflow using MidJourney.

Given GenTune's strength in both efficiency and quality, many participants (P2–5, P7, P9, P11–18, P21) viewed it as a superior tool for communicating with art directors and clients. As P3 shared, " *On a recent project with a tight deadline, the director needed immediate visuals—MidJourney was too unpredictable, but GenTune offered much better control.* " P11 added, " *I used to spend hours revising after meetings; with GenTune, I can make changes live in front of the client.* "

Overall, these findings underscore GenTune's effectiveness in improving refinement efficiency, precision, and quality, making it a practical and controllable tool for real-world creative workflows.

5.3 A3: Alignment with Expectations and Control

In the within-subject study, participants found that GenTune's refinements significantly aligned with their expectations (Fig.4, Q8; Mean_diff = 1.2, p = 0.002) and effectively captured their intended refinement (Fig.4, Q10; Mean_diff = 1.45, p < 0.001). " *The result wasn't exactly what I imagined, but it evolved in a different direction—often exceeding my expectations* " (P9).

Participants rated GenTune as significantly more controllable than the baseline (Fig.4, Q9; Mean_diff = 1, p = 0.003), with 85% preferring its controllability over their previous approach (Fig.5, Q8; Mean = 5.55, p < 0.001). As P14 shared, " *I love the labeling—it gives a sense of control. Even if the result isn't exactly what I imagined, it refines the right area, so the outcome still feels expected.* " This was echoed in open-ended responses: all participants identified label-based selection and editing as GenTune's most helpful feature, with many (P2, P4, P6 – 10, P12 – 13, P16 – 21) attributing their sense of control to it. As P4 put it, " *I finally felt like I was controlling the AI—other tools feel completely random.* "

Participants were significantly more satisfied with GenTune compared to the baseline (Fig.4, Q11; Mean_diff = 1.3, p < 0.001). 90% preferred GenTune for refining AI-generated images (Fig.5, Q9; Mean = 5.6, p < 0.001), 80% found it more satisfying to use (Fig.5, Q10; Mean = 5.55, p < 0.001) than their previous workflows. All participants expressed interest in using GenTune in future commercial projects. Several noted that it increased their trust in generative design tools (P3–4, P6, P11, P13–16, P21). As P6 stated, " *I can clearly see what the AI will generate, which greatly boosts my confidence in using AI.* "

Participants also favored GenTune for its simplicity and ease of control. This aligns with participants' preference for GenTune in supporting creative exploration (Fig.5, Q7; Mean = 5.55, p < 0.001), with 85% favoring it over prior workflows. As P17 noted, " *Compared to tools like ComfyUI or Stable Diffusion—where you have to adjust CFG, weights, models, and parameters—GenTune makes it easy. Designers can just focus on the image and the area they want to refine, and it generates exactly what they need.* " This reflects key principles for tools that support creative thinking [40].

In summary, participants found GenTune to be more controllable and effective for refining AI-generated images. Its intuitive label-based workflow reduced trial and error, while boosting satisfaction, trust, and willingness to adopt it in real-world processes.

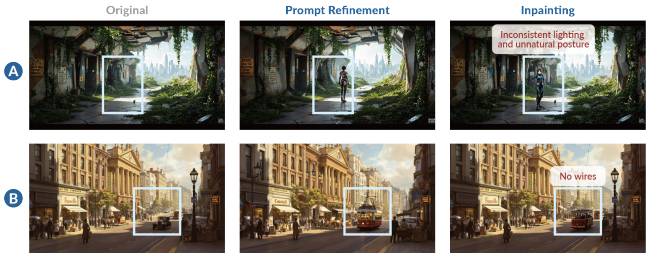

Figure 6: Comparison between semantic-guided prompt refinement with controlled seed (middle) and semantic-guided inpainting (right). In these cases, seed-based refinement better preserves overall image aesthetic coherence.

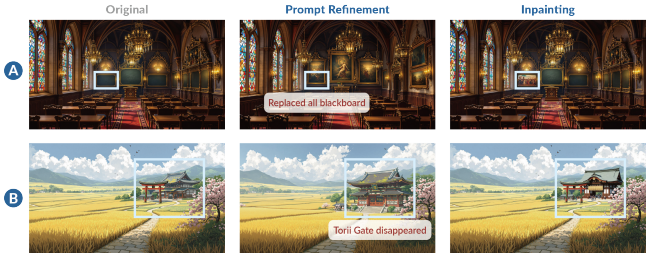

Figure 7: Comparison between semantic-guided prompt refinement with controlled seed (middle) and semantic-guided inpainting (right). In this case, inpainting method provides more precise control.

6 DISCUSSION, LIMITATIONS, AND FUTURE WORK

6.1 Adapting Generation Approaches to Designers' Needs

In both our studies, most participants showed a preference for prompt refining over inpainting for local refinement—even though the former sometimes introduced minor changes in structure or detail. In the within-subject study, among 80 total refinement actions , 47 involved prompt refining, while 33 used inpainting.

Participants attributed their preference to improvements in quality and consistency. Figure 6 shows two comparisons from our open-ended study, where each image pair was generated using mixed mode—once with inpainting and once with prompt refinement. " *When adding people, the lighting and posture look more consistent and natural with the seed* " (Fig.6A, P17). The participants also noted that " *prompt refinement tends to make reasonable changes and offer more possibilities* " (P3). As P16 shared, " *When I added a tram, GenTune also added power lines—I hadn't even thought of that* " (Fig. 6-B, P16). Most designers were comfortable with the slight variations introduced by prompt refining. This reflects a broader value in environment design: aesthetic coherence and scene plausibility often take precedence over strict pixel-level consistency. This aligns with " The Concept of Coherence in Art " [4], which highlights "fittingness" and unity as central to aesthetic experience, even amid minor inconsistencies.

However, despite offering greater consistency, prompt refinement has notable limitations. For example, P4 attempted to replace a single blackboard with a painting (Fig.7-A), but semantic-guided refinement replaced all instances associated with the " blackboard " label, while inpainting made the change more precisely. In Figure 7-B, P12 replaced a shrine with an older version, preserving the structure, but the torii gate was lost. In contrast, inpainting retained the torii and delivered the expected result with higher fidelity.

To support diverse refinement needs, GenTune adopts a mixed mode design, generating two results from each method—semantic-guided prompt refinement and inpainting. This approach offers two key advantages: it helps designers when they are uncertain about the scale of modification, and it enables side-by-side comparison, allowing them to choose the result that aligns with their intent while balancing the strengths and limitations of each method.

6.2 Toward Controllability, Trust, Transparency with Traceable Prompt

Traditional creative workflows are rapidly evolving, with industry professionals increasingly integrating GenAI tools into their pipelines [3, 28, 44]. To meet tight deadlines, designers often rely on LLMs for prompt tuning and visual refinement, intensifying human-AI collaboration [24, 25, 46]. This shift toward automation is accelerated by advances in multi-agent systems and MLLMs [14, 17, 42], which interpret minimal input and autonomously execute complex tasks [20, 33, 39, 52]. A widely discussed example is the image editing tools in Gemini and ChatGPT, where users simply upload an image and provide short instructions, with the system chaining multiple models to complete the task. While these pipelines can produce polished results, their opaque, automated nature limits the user's understanding of the system capabilities and restricts the ability of designers to express intent, clearly articulate their needs, and refine the outputs, a challenge echoed in previous work on human-centered GenAI [15].

Previous approaches have supported the expression of creative intent through multimodal prompting [35] or structured prompt languages [47], and improved control via spatial conditioning, such as region sketches [27] or pen-based tools for color and texture [36]. These methods focus largely on the input stage. In contrast, GenTune takes a post-generation tracing approach: it begins with a brainstorming module that expands simple input into rich prompts, then uses traceable prompts to show how specific components influence different image elements. This trade-off is tailored for environment designers in pre-production workflows, who must rapidly explore visual directions, produce multiple variations, and iteratively refine concepts for stakeholder communication. In such fast-paced settings, manually crafting each design from scratch or repeatedly rewriting prompts is impractical.

Our findings demonstrate that GenTune's traceable prompt significantly enhances designers' sense of control, transparency, trust, and alignment with creative intent. By supporting visual-to-text traceability, GenTune helps environment designers better steer the generation process while preserving their creative autonomy. This approach aligns with core principles of human-centered AI [5, 48, 58]. Future work should further explore integrating domain-specific practices into generative systems, empowering designers across creative fields to effectively leverage advanced AI capabilities while maintaining creative autonomy [19].

6.3 Limitations and Future Work

Refinement order. While combining inpainting and prompt refinement offers flexibility and improves the chances of achieving a satisfactory result, it also introduces a key limitation: once inpainting is applied, subsequent prompt refinement may regenerate the entire image and overwrite earlier edits. This remains a significant challenge, as designers may struggle to anticipate which method is best suited for a given modification. As P3 noted, " *You need to plan ahead—once you inpaint, it's difficult to go back and change the overall image.* " GenTune addresses this through a layered system that tracks changes across methods and allows designers to revert to previous versions. Future work could enhance this by introducing a rapid preview system, enabling users to compare refinement outcomes quickly.

Instability of semantic-guided prompt refinement. Some participants noted the instability of prompt refinement. As P21 remarked, " *Sometimes it accurately modifies only the part I intended, but other times, elements I previously edited disappear.* " This instability stems from the limitations of T2I models, which often struggle to maintain structural and spatial consistency when prompt intent shifts—even with a fixed seed [7, 13]. Recent work has begun addressing these issues through improved prompt refinement [31], enhanced spatial understanding [13], and multi-view consistency [22]. The development of spatially conditioned control models [60] also opens new possibilities for more stable and consistent refinements. Future work could integrate these control strategies to strengthen the refinement process further.

Label accuracy. The effectiveness of GenTune's refinement relies heavily on accurate label selection. If the correct label is missing or unclear, the refinement may not reflect the designer's intent. Label inaccuracies typically arise from: (1) hallucinations in the T2I model, where elements appear visually but are not captured in prompt-derived labels, and (2) an overabundance of similar or ambiguous labels, making it hard to identify the right one. Beyond improving classification and T2I accuracy, future versions of GenTune could support reverse mapping—highlighting all regions linked to a selected label—to help designers better anticipate which areas will be affected.

7 CONCLUSION

We present GenTune, a human-centered generative AI system that enhances the refinement workflow in environment design. Through Traceable prompts and Semantic-guided refinement, GenTune helps designers better understand prompt–image relationships while enabling more precise and consistent edits. We evaluated GenTune in a summative study with 20 environment designers in an experiment within the subject and an open-ended design task. The results show that GenTune significantly improved prompt image comprehension, refinement quality, and efficiency, and the sense of control of users, giving a clear preference to existing workflows. A field study in two design studios further demonstrated GenTune's potential to improve refinement efficiency and creative communication in real-world production settings.

References

[1] 3dtotal Publishing. 2018. *The Ultimate Concept Art Career Guide*. 3dtotal Publishing.

[2] M3DS Academy. 2024. Environment Designer. https://www.artstation.com/blogs/m3dsacademy/zXXz6/environment-designer.

[3] Akash Anant Alegaonkar and Mukta Aditya Avachat-Shirke. 2023. Is Artificial Intelligence Killing Artistic Skills in Designers?. In *International Conference on Emerging Trends in Design & Arts*, Vol. 4. 109–115.

[4] L Aschenbrenner. 2012. *The concept of coherence in art*. Springer Science & Business Media.

[5] Jan Auernhammer. 2020. Human-centered AI: The role of Human-centered Design Research in the development of AI. (2020).

[6] Eleonora Vilgia Putri Beyan, Anastasya Gisela Cinintya Rossy, et al. 2023. A review of AI image generator: influences, challenges, and future prospects for architectural field. *Journal of Artificial Intelligence in Architecture* 2, 1 (2023), 53–65.

[7] Zenab Bosheah and Vilmos Bilicki. 2025. Challenges in Generating Accurate Text in Images: A Benchmark for Text-to-Image Models on Specialized Content. *Applied Sciences* 15, 5 (2025), 2274.

[8] 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*. 1–14.

[9] Ross Brisco, Laura Hay, and Sam Dhami. 2023. Exploring the role of text-to-image AI in concept generation. *Proceedings of the Design Society* 3 (2023), 1835–1844.

[10] Alice Cai, Steven R Rick, Jennifer L Heyman, Yanxia Zhang, Alexandre Filipowicz, Matthew Hong, Matt Klenk, and Thomas Malone. 2023. DesignAID: Using generative AI and semantic diversity for design inspiration. In *Proceedings of The ACM Collective Intelligence Conference*. 1–11.

[11] Marinela Capanu, Gregory A Jones, and Ronald H Randles. 2006. Testing for preference using a sum of Wilcoxon signed rank statistics. *Computational statistics & data analysis* 51, 2 (2006), 793–796.

[12] cgspectrum. 2024. Environment Designer. https://www.cgspectrum.com/career-pathways/environment-designer.

[13] Agneet Chatterjee, Gabriela Ben Melech Stan, Estelle Aflalo, Sayak Paul, Dhruba Ghosh, Tejas Gokhale, Ludwig Schmidt, Hannaneh Hajishirzi, Vasudev Lal, Chitta Baral, et al. 2024. Getting it right: Improving spatial consistency in text-to-image models. In *European Conference on Computer Vision*. Springer, 204–222.

[14] Qiguang Chen, Libo Qin, Jinhao Liu, Dengyun Peng, Jiannan Guan, Peng Wang, Mengkang Hu, Yuhang Zhou, Te Gao, and Wangxiang Che. 2025. Towards reasoning era: A survey of long chain-of-thought for reasoning large language models. *arXiv preprint arXiv:2503.09567* (2025).

[15] Xiang'Anthony' Chen, Jeff Burke, Ruofei Du, Matthew K Hong, Jennifer Jacobs, Philippe Laban, Dingzeyu Li, Nanyun Peng, Karl DD Willis, Chien-Sheng Wu, et al. 2023. Next steps for human-centered generative AI: A technical perspective. *arXiv preprint arXiv:2306.15774* (2023).

[16] ComfyUI Contributors. 2023. ComfyUI: A powerful and modular Stable Diffusion GUI and backend. https://github.com/comfyanonymous/ComfyUI.

[17] Yuhao Dong, Zuyan Liu, Hai-Long Sun, Jingkang Yang, Winston Hu, Yongming Rao, and Ziwei Liu. 2024. Insight-v: Exploring long-chain visual reasoning with multimodal large language models. *arXiv preprint arXiv:2411.14432* (2024).

[18] Lauren du Plessis. 2022. What Is Game Environment Design and How to Get Started? https://www.domestika.org/en/blog/10804-what-is-game-environment-design-and-how-to-get-started.

[19] Thomas F Eisenmann, Andres Karjus, Mar Canet Sola, Levin Brinkmann, Bramantyo Ibrahim Supriyatno, and Iyad Rahwan. 2025. Expertise elevates AI usage: experimental evidence comparing laypeople and professional artists. *arXiv preprint arXiv:2501.12374* (2025).

[20] Difei Gao, Lei Ji, Zechen Bai, Mingyu Ouyang, Peiran Li, Dongxing Mao, Qinchen Wu, Weichen Zhang, Peiyi Wang, Xiangwu Guo, et al. 2024. Assistgui: Task-oriented pc graphical user interface automation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13289–13298.

[21] Yuying Ge, Sijie Zhao, Chen Li, Yixiao Ge, and Ying Shan. 2024. Seed-data-edit technical report: A hybrid dataset for instructional image editing. *arXiv preprint arXiv:2405.04007* (2024).

[22] Lukas Höllein, Aljaž Božič, Norman Müller, David Novotny, Hung-Yu Tseng, Christian Richardt, Michael Zollhöfer, and Matthias Nießner. 2024. Viewdiff: 3d-consistent image generation with text-to-image models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5043–5052.

[23] Minbin Huang, Yanxin Long, Xinchi Deng, Ruihang Chu, Jiangfeng Xiong, Xiaodan Liang, Hong Cheng, Qinglin Lu, and Wei Liu. 2024. Dialoggen: Multi-modal interactive dialogue system for multi-turn text-to-image generation. *arXiv preprint arXiv:2403.08857* (2024).

[24] Harry H Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. AI Art and its Impact on Artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 363–374.

[25] Reishiro Kawakami and Sukrit Venkatagiri. 2024. The Impact of Generative AI on Artists. In *Proceedings of the 16th Conference on Creativity & Cognition*. 79–82.

[26] Eliott J. Lilly. 2015. *Big Bad World of Concept Art for Video Games: An Insider's Guide for Students*. Design Studio Press.

[27] Haichuan Lin, Yilin Ye, Jiazhi Xia, and Wei Zeng. 2025. SketchFlex: Facilitating Spatial-Semantic Coherence in Text-to-Image Generation with Region-Based Sketches. *arXiv preprint arXiv:2502.07556* (2025).

[28] LING Long, CHEN Xinyi, WEN Ruoyu, LI Toby Jia-Jun, and LC Ray. 2024. Sketchar: Supporting Character Design and Illustration Prototyping Using Generative AI. *Proceedings of the ACM on Human-Computer Interaction* 8, CHI PLAY (2024), 337.

[29] Gianmarco Longo, Deborah Middleton, and Silvia Albano. 2024. Elaborating a framework that is able to structure and evaluate design workflow and composition of Generative AI Visualizations.. In *IHIET-AI 2024: 11th International Conference on Human Interaction & Emerging Technologies: Artificial Intelligence & Future Applications*. AHFE International Open Access.

[30] Sebastian Lubos, Thi Ngoc Trang Tran, Alexander Felfernig, Seda Polat Erdeniz, and Viet-Man Le. 2024. LLM-generated Explanations for Recommender Systems. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. 276–285.

[31] Oscar Mañas, Pietro Astolfi, Melissa Hall, Candace Ross, Jack Urbanek, Adina Williams, Aishwarya Agrawal, Adriana Romero-Soriano, and Michal Drozdzal. 2024. Improving text-to-image consistency via automatic prompt optimization. *arXiv preprint arXiv:2403.17804* (2024).

[32] Adrian Marc. 2023. *The Random Guidebook of Concept Designers : Tips and Tricks* (1st ed.). JOLUA.

[33] Boye Niu, Yiliao Song, Kai Lian, Yifan Shen, Yu Yao, Kun Zhang, and Tongliang Liu. 2025. Flow: A Modular Approach to Automated Agentic Workflow Generation. *arXiv preprint arXiv:2501.07834* (2025).

[34] Srishti Palani, David Ledo, George Fitzmaurice, and Fraser Anderson. 2022. " I don ' t want to feel like I ' m working in a 1960s factory " : The Practitioner Perspective on Creativity Support Tool Adoption. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[35] Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2024. DesignPrompt: Using Multimodal Interaction for Design Exploration with Generative AI. (2024), 804–818. https://doi.org/10.1145/3643834.3661588

[36] Xiaohan Peng, Janin Koch, and Wendy E Mackay. 2025. FusAIn: Composing Generative AI Visual Prompts Using Pen-based Interaction. (2025).

[37] Manisha Pise, Naveen Yadgiri, Preksha Gaikwad, Yashika Dusawar, and Prathamesh Nandanwar. 2024. AI Image Generator. *networks (GANs)* 4, 4 (2024).

[38] CB Pronin, AA Podberezkin, and AM Borzenkov. 2024. Evaluating Consistency of Image Generation Models with Vector Similarity. In *2024 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED)*. IEEE, 1–4.

[39] Xihe Qiu, Haoyu Wang, Xiaoyu Tan, Chao Qu, Yujie Xiong, Yuan Cheng, Yinghui Xu, Wei Chu, and Yuan Qi. 2024. Towards Collaborative Intelligence: Propagating Intentions and Reasoning for Multi-Agent Coordination with Large Language Models. *arXiv preprint arXiv:2407.12532* (2024).

[40] Mitchel Resnick, Brad Myers, Kumiyo Nakakoji, Ben Shneiderman, Randy Pausch, Ted Selker, and Mike Eisenberg. 2005. Design principles for tools to support creative thinking. (2005).

[41] Paula K Roberson, SJ Shema, DJ Mundfrom, and TM Holmes. 1995. Analysis of paired Likert data: how to evaluate change and preference questions. *Family medicine* 27, 10 (1995), 671–675.

[42] Manish Sanwal. 2025. Layered Chain-of-Thought Prompting for Multi-Agent LLM Systems: A Comprehensive Approach to Explainable Large Language Models. *arXiv preprint arXiv:2501.18645* (2025).

[43] Arvind Satyanarayan, Bongshin Lee, Donghao Ren, Jeffrey Heer, John Stasko, John Thompson, Matthew Brehmer, and Zhicheng Liu. 2019. Critical reflections on visualization authoring systems. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 461–471.

[44] Ojas D. Sawant. 2024. *Visual Storytelling with Generative AI: A Practical Handbook for modern Filmmakers and Content Creators*. Independently published.

[45] Jesse Schell. 2008. *The Art of Game Design: A Book of Lenses*. CRC Press.

[46] Jingyu Shi, Rahul Jain, Runlin Duan, and Karthik Ramani. 2023. Understanding Generative AI in Art: An Interview Study with Artists on G-AI from an HCI Perspective. *arXiv preprint arXiv:2310.13149* (2023).

[47] Xinyu Shi, Yinghou Wang, Ryan Rossi, and Jian Zhao. 2025. Brickify: Enabling Expressive Design Intent Specification through Direct Manipulation on Design Tokens. *arXiv preprint arXiv:2502.21219* (2025).

[48] Ben Shneiderman. 2022. *Human-centered AI*. Oxford University Press.

[49] 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*. 1–19.

[50] SM Taheri and Gholamreza Hesamian. 2013. A generalization of the Wilcoxon signed-rank test and its applications. *Statistical Papers* 54 (2013), 457–470.

[51] Yuying Tang, Mariana Ciancia, Zhigang Wang, and Ze Gao. 2024. What's Next? Exploring Utilization, Challenges, and Future Directions of AI-Generated Image Tools in Graphic Design. *arXiv preprint arXiv:2406.13436* (2024).
[52] Chenyu Wang, Weixin Luo, Qianyu Chen, Haonan Mai, Jindi Guo, Sixun Dong, Zhengxin Li, Lin Ma, Shenghua Gao, et al. 2024. Mllm-tool: A multimodal large language model for tool agent learning. *arXiv preprint arXiv:2401.10727* (2024).
[53] Wen-Fan Wang, Chien-Ting Lu, Nil Ponsa Campanyà, Bing-Yu Chen, and Mike Y Chen. 2025. AIdeation: Designing a Human-AI Collaborative Ideation System for Concept Designers. *arXiv preprint arXiv:2502.14747* (2025).
[54] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. 2023. Reprompt: Automatic prompt editing to refine ai-generative art towards precise expressions. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–29.
[55] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. , Article 185 (2024), 21 pages. https://doi.org/10.1145/3613904.3642803
[56] Jingxuan Wei, Shiyu Wu, Xin Jiang, and Yequan Wang. 2023. Dialogpaint: A dialog-based image editing model. *arXiv preprint arXiv:2303.10073* (2023).
[57] Robert F Woolson. 2005. Wilcoxon signed-rank test. *Encyclopedia of biostatistics* 8 (2005).
[58] Wei Xu, Marvin J Dainoff, Liezhong Ge, and Zaifeng Gao. 2023. Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI. *International Journal of Human–Computer Interaction* 39, 3 (2023), 494–518.
[59] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. 2018. Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5505–5514.
[60] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*. 3836–3847.