RESEARCH-ARTICLE

GenTune: Toward Traceable Prompts to Improve Controllability of Image Refinement in Environment Design

WEN-FAN WANG, National Taiwan University, Taipei, Taiwan

TING-YING LEE, National Taiwan University, Taipei, Taiwan

CHIEN-TING LU, National Taiwan University, Taipei, Taiwan

CHE-WEI HSU, National Taiwan University, Taipei, Taiwan

NIL PONSA I CAMPANYÀ, National Taiwan University, Taipei, Taiwan

YU CHEN, National Taiwan University, Taipei, Taiwan

View all

Published: 27 September 2025

UIST '25: The 38th Annual ACM Symposium on User Interface Software and Technology
September 28 - October 1, 2025
Busan, Republic of Korea

Conference Sponsors:
SIGCHI
SIGGRAPH

GenTune: Toward Traceable Prompts to Improve Controllability of Image Refinement in Environment Design

Wen-Fan Wang*
National Taiwan University
Taipei, Taiwan
vann@cmlab.csie.ntu.edu.tw

Ting-Ying Lee*
National Taiwan University
Taipei, Taiwan
tylee@cmlab.csie.ntu.edu.tw

Chien-Ting Lu
National Taiwan University
Taipei, Taiwan
b09902109@csie.ntu.edu.tw

Che-Wei Hsu
National Taiwan University
Taipei, Taiwan
yaweihsu1234@gmail.com

Nil Ponsa i Campanyà
National Taiwan University
Taipei, Taiwan
r12944063@csie.ntu.edu.tw

Yu Chen
National Taiwan University
Taipei, Taiwan
r11922026@ntu.edu.tw

Mike Y. Chen
National Taiwan University
Taipei, Taiwan
mikechen@csie.ntu.edu.tw

Bing-Yu Chen
National Taiwan University
Taipei, Taiwan
robin@ntu.edu.tw

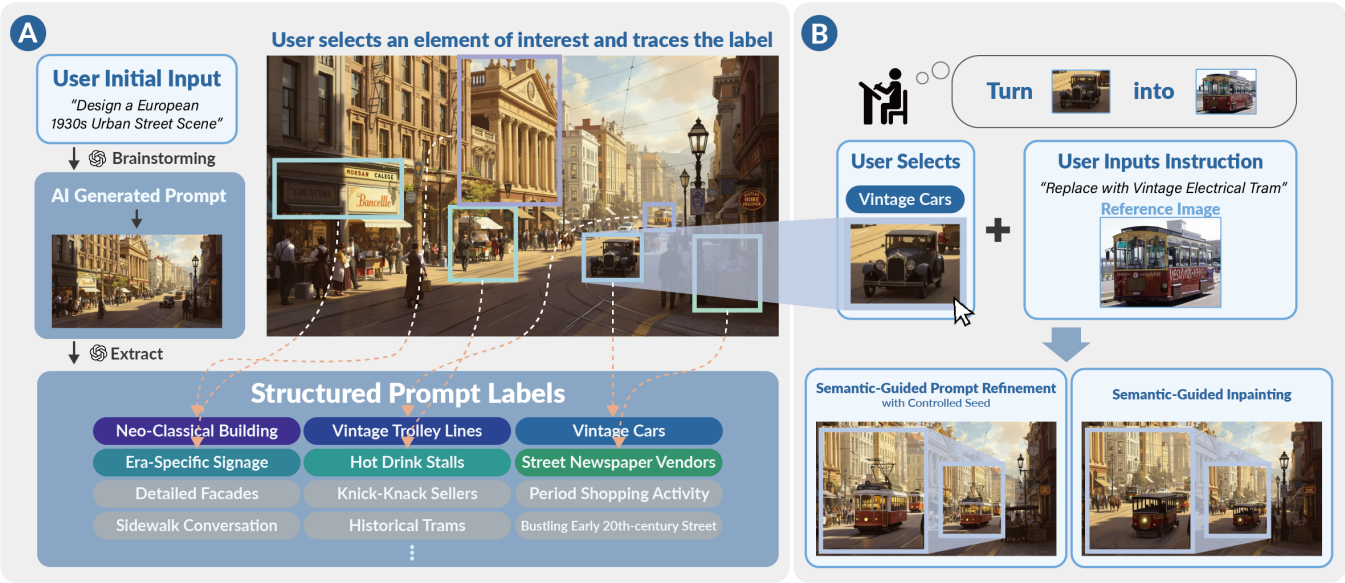

Figure 1: GenTune, a human-centered generative AI system with traceable prompts for controllable image refinement in environment design. (A) The process begins with a user's initial input-"Design a European 1930s Urban Street Scene"-which is expanded by a Brainstorming LLM into a structured prompt for 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.

*Both authors contributed equally as first author.

UIST '25, Busan, Republic of Korea

Abstract

Environment designers in the entertainment industry create imaginative 2D and 3D scenes for games, films, and television, requiring both fine-grained control of specific details and consistent global coherence. Designers have increasingly integrated generative AI

ACM ISBN 979-8-4007-2037-6/25/09
https://doi.org/10.1145/3746059.3747774

into their workflows, often relying on large language models (LLMs) to expand user prompts for text-to-image generation, then iteratively refining those prompts and applying inpainting. However, our formative study with 10 designers surfaced two key challenges: (1) the lengthy LLM-generated prompts make it difficult to understand and isolate the keywords that must be revised for specific visual elements; and (2) while inpainting supports localized edits, it can struggle with global consistency and correctness. Based on these insights, we present GenTune, an approach that enhances human–AI collaboration by clarifying how AI-generated prompts map to image content. Our GenTune system lets designers select any element in a generated image, trace it back to the corresponding prompt labels, and revise those labels to guide precise yet globally consistent image refinement. In a summative study with 20 designers, GenTune significantly improved prompt-image comprehension, refinement quality and efficiency, and overall satisfaction (all $p < .01$) compared to current practice. A follow-up field study with two studios further demonstrated its effectiveness in real-world settings.

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:
Wen-Fan Wang, Ting-Ying Lee, Chien-Ting Lu, Che-Wei Hsu, Nil Ponsa i Campanyà, Yu Chen, Mike Y. Chen, and Bing-Yu Chen. 2025. GenTune: Toward Traceable Prompts to Improve Controllability of Image Refinement in Environment Design. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25), September 28–October 01, 2025, Busan, Republic of Korea.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3746059.3747774

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, 70, 84, 102]. 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 [102]. These designs often serve as blueprints for modeling and VFX teams, or in smaller or stylized productions, are used directly in final scenes [20, 35]. 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, 70, 84].

With the rise of generative AI (GenAI) tools, environment designers increasingly integrate them into their workflows [77, 78, 101, 122]. They collaborate with multimodal large language models (MLLMs) to craft and refine prompts, which serve as input for text-to-image (T2I) models. GenAI is used across ideation, inspiration, client communication, and even production-level outputs [38, 90, 122]. 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 [13, 124, 126] and multimodal models for rapid ideation [18, 89, 122].

However, as designers move to refinement, they often need to modify or build upon initial AI-generated outputs. 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 structural coherence, and their automated nature limits designer control and restricts expressive intent. As a result, designers struggle with precision in the refinement process, often resorting to time-consuming trial-and-error cycles [10, 15, 112, 122]. In this work, we take a human-centered approach to support understanding and refinement of both AI-generated intermediate prompts and final outputs, tailored to the specific needs of environment designers in the pre-production process.

To understand the challenges environment designers face, we conducted a formative study with 10 participants. Through workflow analysis and in-depth interviews, we identified two core challenges: a lack of understanding of how intermediate 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. Designers often rely on large language models (LLMs) to generate lengthy, detailed prompts that guide the initial image generation process. They begin with global refinement, adjusting the entire image, including layout or composition. However, for local refinement, they frequently struggle to trace how specific parts of the prompt influence the visual elements, making refinement a trial-and-error process. While inpainting [135] enables targeted edits, it often introduces inconsistencies in lighting, style, or context. More technical solutions, such as ComfyUI [31] with ControlNet [143], offer finer control but are too complex and misaligned with designers' workflows.

To address this, 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 [18, 122] by incorporating an MLLM module that expands simple user inputs into detailed prompts. 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 traced from the expanded prompt used to generate the image (Fig. 1-A). The full prompt segment associated with the label is also displayed, helping designers understand the prompt-image relationship.
(2) **Semantic-Guided Refinement:** GenTune allows designers to precisely refine the visual elements based on a selected area of interest (Fig. 1-B). The system supports three refinement modes: refining the visual 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.

We conducted a summative study with 20 participants (15 professionals, 5 design students), all experienced with GenAI tools. Across both a controlled experiment and an open-ended task using their own projects, participants rated GenTune significantly higher than a baseline that lacks GenTune's two key modules in prompt–image understanding ($p = 0.003$), refinement effectiveness ($p = 0.002$), output quality ($p = 0.003$), and overall satisfaction ($p < 0.001$). Compared to their typical workflows, they also preferred GenTune for satisfaction, efficiency, quality, and creativity support ($p < 0.001$ for all). Following the summative study, we conducted a field study with two design studios to evaluate GenTune in real-world professional settings. Designers integrated GenTune into their ongoing commercial projects for three days. All participants reported improved efficiency, higher output quality, and a significant reduction in time spent communicating design intent to clients and directors.

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.
- A HCI paradigm for traceable, element-level control, enhancing understanding and controllability in human–AI collaboration.
- 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, (2) an open-ended task using production projects, and (3) a field deployment in two design studios.

2 RELATED WORK

2.1 Generative AI in Creative Workflow

The creative process involves both divergent and convergent thinking. Designers begin by exploring a wide range of possibilities to generate diverse ideas [61, 129, 130, 144], and iteratively refine selected concepts through evaluation cycles [56, 60, 107]. This process often requires a non-linear, iterative process between exploration and refinement [4, 28, 43].

Previous work has leveraged GenAI to support divergent stages of creativity, aiding visual exploration and early ideation. These systems help users generate various outputs [75], explore unexpected directions [64], and have been applied in domains such as fashion [57], architecture [110], and 3D CAD modeling [76]. For visual exploration, some tools streamline prompt crafting for faster feedback [30, 74], while others support conceptual expansion and recombination. For example, CreativeConnect [27], PopBlends [120], and GenQuery [106] help designers decompose and blend concepts to generate novel ideas. DesignAID [18] uses LLMs to expand design space, and AIdeation [122] further enables flexible recombination of retrieved visual references and generated content.

Recent work has integrated generative AI into the creative field to enhance domain-specific workflows and communication throughout the design process. For example, MemoVis [22] translates textual feedback into visual references to reduce misinterpretation in 3D modeling, while RoomDreaming [121] enables collaborative spatial exploration through photorealistic interiors. PlantoGraphy [54] visualizes abstract prompts for iterative landscape design, and Keyframer [114] offers a natural language interface for motion design. Other applications span video co-creation [55], creative writing [44], and spatial design [117]. These tools demonstrate the potential of GenAI to accelerate exploration and improve communication.

Unlike most prior systems that focus on divergent idea exploration, GenTune extends these capabilities into the refinement stage—supporting convergent processes and addressing challenges such as maintaining visual consistency, a key challenge in current environment designers' GenAI workflows.

2.2 Addressing Understanding in Human-AI Design Collaboration

Understanding and effectively interacting with GenAI remains a challenge, especially for non-experts [62]. Despite advances in T2I models, users often struggle to clearly express their intention and achieve the desired results [12, 139]. A key challenge lies in the unintuitive nature of prompt crafting, which requires careful planning and remains difficult for non-expert users [137, 138]. While many systems aim to optimize prompts [5, 118], manual prompt creation is often still necessary, especially for environment designers, who heavily rely on commercial tools like MidJourney[1]. Although some artists and designers use LLMs to assist in prompt generation [68, 73], a gap remains in users' understanding of T2I models. They often struggle to connect prompts with generated outputs [85], resulting in limited control and reliance on trial and error [47].

To address these challenges, Explainable AI (XAI) approaches aim to make the generation process more transparent and interpretable [9, 36, 123]. However, much of this work remains system-centered, focusing on algorithmic transparency over user understanding [62]. Recent HCI research has begun to bridge this gap with human-centered AI approaches [37, 69, 115], emphasizing user engagement and dynamic interaction [93, 113]. This shift prioritizes interpretability as a means to improve user experience and support human-AI collaboration.

For example, From Text to Pixels [41] uses visual and sensitivity-based explanations to show how text prompts influence outputs in T2I models. Other systems support fine-grained understanding through different modalities: AutoSpark [23] allows detailed comparisons to improve text-image relevance; XCreation [134] offers a graph-based interface for intuitive control in cross-modal story creation; and ProtoDreamer [141] supports physical prototyping with AI feedback on design evolution. Interactive systems in music [79], translation [32], and crowd ideation [125] likewise enhance user understanding by exposing AI reasoning and supporting more directed interaction.

Whereas prior works adopt diverse strategies to enhance user understanding of the generative process, GenTune's focus on prompt-to-image element mapping is action-oriented: traceability is not the end goal, but a mechanism to enable direct, element-centric manipulation and refinement.

[1] MidJourney, https://www.midjourney.com/

2.3 Refinement in Generative Image Workflows

While generative AI has been widely studied for its role in the divergent stages of creativity, such as ideation and iterative design [87, 108], fewer studies have examined its potential to support convergent processes, as highlighted in recent research [26, 96].

Recent work has supported image refinement through interactive and automated tools. RePrompt [124] and Promptify [13] enable iterative prompt editing, with RePrompt using CLIP-based scoring for alignment. PromptCharm [126] and PromptMagician [42] offer interfaces for exploring and editing prompt variations, with PromptCharm additionally suggesting terms based on user intent. PromptMap [3] enables semantic navigation across prompt spaces, while DreamSheets [6] visualizes prompt exploration to aid discovery and refinement. Fermat [103] provides multimodal control, supporting both exploratory and refinement phases. These systems alleviate the trial-and-error burden of prompt crafting by directly manipulating prompts. In contrast, GenTune utilizes LLMs to generate structured prompts from simple input and enables refinement post-generation, thereby bypassing the need for manual prompt crafting.

Beyond prompt-based refinement, recent work has explored image-based techniques—mainly built on Stable Diffusion [97, 98]—to enhance control over generative outputs. LoRA [52] enables lightweight style tuning, while InstructPix2Pix [16] allows edits via natural language. Inpainting methods [81, 136] support selective region regeneration. ControlNet-based tools [143], using inputs like edges, depth, or segmentation maps, provide structural conditioning and are widely adopted in platforms like ComfyUI [133]. However, these methods often come with steep learning curves [140] and may misalign with creative workflows. Seed-based generation offers an alternative, where consistent seeds and detailed prompts yield stable results [91, 131], but this typically demands precise prompt engineering [46, 67].

Beyond these methods, multimodal interfaces have been explored to provide more expressive and localized control in generative image workflows. A common approach is sketch-based interaction: SketchFlex [72] refines rough sketches into structured anchors for spatial-semantic coherence, while Inkspire [71] supports analogical sketching for early-stage ideation. PromptPaint [29] and Exploring Visual Prompts [88] use scribbles, annotations, and brushes to guide generation. DesignPrompt [89] combines images, colors, and text in a moodboard-style prompt interface. Other systems support human–AI co-creation via multimodal control: GANzilla [39] spans both exploratory and refinement phases, while GANravel [40] enables user-driven direction disentanglement for iterative GAN editing.

More recently, MLLMs like ChatGPT-4o and Gemini 2.0 Flash[2] enable image editing through dialogue-based interactions, offering prompt enhancement and structural control. While promising, they still fall short in meeting the specific refinement needs of environment design, as discussed later.

While prior work focuses on diverse refinement approaches, few address multi-step, LLM-assisted input generation, where controlling intermediate outputs is critical. GenTune introduces traceable, element-level control to improve understanding and controllability in future MLLM-assisted workflows.

[2]Gemini, https://developers.googleblog.com/en/experiment-with-gemini-20-flash-native-image-generation/

3 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.

3.1 Participants

We recruited 10 participants, including 5 professional environment designers from the game (P1, P4, P6) and animation (P3, P5) industries (3–8 YoE, mean = 4.8), and 5 design students (P2, P7–P10) with at least six months of intensive GenAI experience. Participants were recruited through personal referrals within the local art community and by directly contacting studios by email to request collaboration. Participants were compensated 15 $USD. Demographics, including study participation, years of experience, and commonly used AI tools, are shown in Table 1.

3.2 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. Each session lasted 1–1.5 hours.

3.3 Findings

3.3.1 Data analysis. We conducted a thematic analysis of summarized and transcribed interviews. An author with professional environment design experience developed the initial coding framework, which was refined with feedback from two art directors overseeing 21 and 34 concept designers. Thematic analysis was discussed collaboratively among three co-author to ensure consensus.

3.3.2 Roles of GenAI in Environment Design. Designers commonly use GenAI tools for early-stage ideation. They begin by using ML models to expand simple prompts and generate images as visual references, especially when working with unfamiliar design specifications (P2-4, P6-9). As one participant noted, "*It's common that we cannot find specific design references to the topic, but GenAI tools can provide that*" (P4). 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 (P1, P3-6). 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).

3.3.3 Current GenAI Refinement Workflow and Key Challenges. As shown in Table 1, participants used various image generation

tools. Despite those differences, their workflows were similar: after initial generation, they refined the images iteratively from global structures to local details. During the refinement, we found designers often emphasize several elements such as theme, art style, composition, lighting, color, and shot angle.

They usually begin with prompt editing. However, a key challenge was the lack of clear mapping between the LLM-generated prompt and visual elements, making it hard to determine which parts of the prompt control specific image features (P2-4, P6-7, P9-10). "*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; for these, neither manual edits nor LLM assistance are effective. These challenges underscore the inefficiencies of prompt-based refinement, especially in complex environment design.

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 (P1, P3-5, P7-9). "*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 [31], 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 [135]. This involves selecting a region and entering a replacement prompt. However, results were often visually inconsistent—mismatched lighting, style, proportion, or contextual incoherence, such as historical or spatial inconsistency (P1-3, P6-9). "*Inpainting often gives me strange results. I've tried many different prompts, but still didn't get what I wanted*" (P6). Designers emphasized that visual coherence was more important (P1, P3-5, P7-10): "*Although inpainting preserved other areas, the materials were completely wrong. I would rather start over.*" (P7). Others opted for manual editing, but the lack of editable layers made this difficult. "*It's faster to repaint the area from scratch*" (P3). Still, when many similar elements needed changes, manual editing "*just takes too much time*" (P4).

In summary, environment designers struggle to understand how LLM-generated prompts map to visual outputs and lack support for targeted refinement. They prioritize visual coherence over pixel-level accuracy—something current GenAI tools often fail to provide.

3.4 Design Goals

Based on the findings, 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.
- **DG2: Maintaining Coherence in Refinement.** The system should align with designers' priorities by maintaining visual and semantic coherence throughout the iterative refinement process, from global structure to local details.
- **DG3: Element-Centric and Predictable Control Workflow.** The system should support element-centric manipulation that offers precise control and predictability, reducing trial-and-error and enhancing designers' creative exploration.

4 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.

4.1 System Overview

GenTune's image generation system draws inspiration from conversational generation systems [53, 122, 127], which support iterative, dialogue-driven workflows. It features a brainstorming module that expands 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.

4.1.1 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 lets designers select an element in the image to reveal a corresponding label traced from the expanded prompt; the related section in the prompt panel then expands automatically (Fig. 2-C1). Labels are organized in a tree structure, with parent categories expanding when traced, and are derived from the "content" category, which maps to identifiable visual elements.

4.1.2 Semantic-guided refinement. After tracing a label, designers can refine the image by entering natural language instructions in the dialogue box, or providing 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)**. Designers can make broad, whole-image edits using text or reference images, such as changing style, mood, or lighting, without selecting specific regions.
- **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

ID	Age	YoE	Industry	GenAI Tools Used	Formative	Summative	Field
1	34	8	Game	Midjourney, ComfyUI	✓		
2	22	0	Student	ChatGPT (DALL·E)	✓	✓	
3	27	3	Animation	Midjourney	✓	✓	✓
4	27	4	Game	Midjourney, ChatGPT (DALL·E)	✓	✓	✓
5	27	5	Animation	Midjourney	✓	✓	
6	27	4	Game	Stable Diffusion	✓	✓	
7	23	0	Student	Midjourney, Stable Diffusion, ChatGPT (DALL·E), Leonardo	✓	✓	
8	22	0	Student	Midjourney, ChatGPT (DALL·E), Leonardo	✓	✓	
9	22	0	Freelancing, Animation, Student	ChatGPT (DALL·E), ComfyUI	✓	✓	
10	23	0	Animation, Student	Midjourney, Stable Diffusion, ChatGPT (DALL·E)	✓	✓	
11	27	4	Game	Midjourney		✓	
12	39	10	Game	Midjourney, Stable Diffusion		✓	
13	39	13	Game	ChatGPT (DALL·E)		✓	
14	24	1	Animation	Midjourney		✓	
15	30	8	Freelancing, Film/TV	Stable Diffusion		✓	
16	27	5	Freelancing, Film/TV, Animation	Midjourney, Stable Diffusion, ChatGPT (DALL·E)		✓	
17	43	21	Game	Stable Diffusion, ChatGPT (DALL·E), ComfyUI		✓	
18	28	5	Freelancing	Midjourney, ChatGPT (DALL·E)		✓	
19	31	6	Freelancing	Stable Diffusion		✓	
20	28	3	Industrial Design	Midjourney, Firefly, Vizcom		✓	
21	30	7	Animation	ChatGPT (DALL·E), Moonshot		✓	
22	27	3	Animation	Midjourney, Stable Diffusion			✓

Table 1: Demographic Details of Participants Including Age, Industry, Generative AI Tools, and Study Participation

only to the intended element while preserving overall coherence [46, 91]. 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. 13-B). This design choice aligns AI-driven refinement with designer intent by prioritizing conceptual consistency and visual coherence over exact pixel-level accuracy.

- **Semantic-guided inpainting (requires selection)** This option enables more localized edits. Unlike traditional inpainting, 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.

As shown in Figure 3, GenTune supports a progressive refinement workflow that aligns with designers' natural editing process—starting from global adjustments to fine-grained control. Users typically begin with Global Refinement for large-scale changes (e.g. art style), followed by Semantic-Guided Prompt Refinement to adjust specific elements while preserving overall structure, and finally Semantic-Guided Inpainting to make precise, localized edits.

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.

GenTune provides three types of refinement suggestions: (1) global suggestions for broad edits like lighting or content (Fig.2-C2); (2) label-based suggestions for element-specific changes (Fig.2-A3); and (3) expanded suggestions that build on user input to offer detailed, context-aware options. These aid designers, especially when unsure of the next step.

4.2 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. 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.

4.3 Technical Implementation

GenTune uses GPT-4o-2024-08-06[3] as the base language model, and Flux 1.1 Pro Ultra[4] as the image model, for its strong performance with natural language prompts. The average generation time for a single iteration is approximately 30 seconds. The backend and frontend are deployed on a Linux-based personal computer with

[3]GPT-4o, https://platform.openai.com/docs/models/gpt-4o
[4]Flux 1.1, https://blackforestlabs.ai/ultra-home/

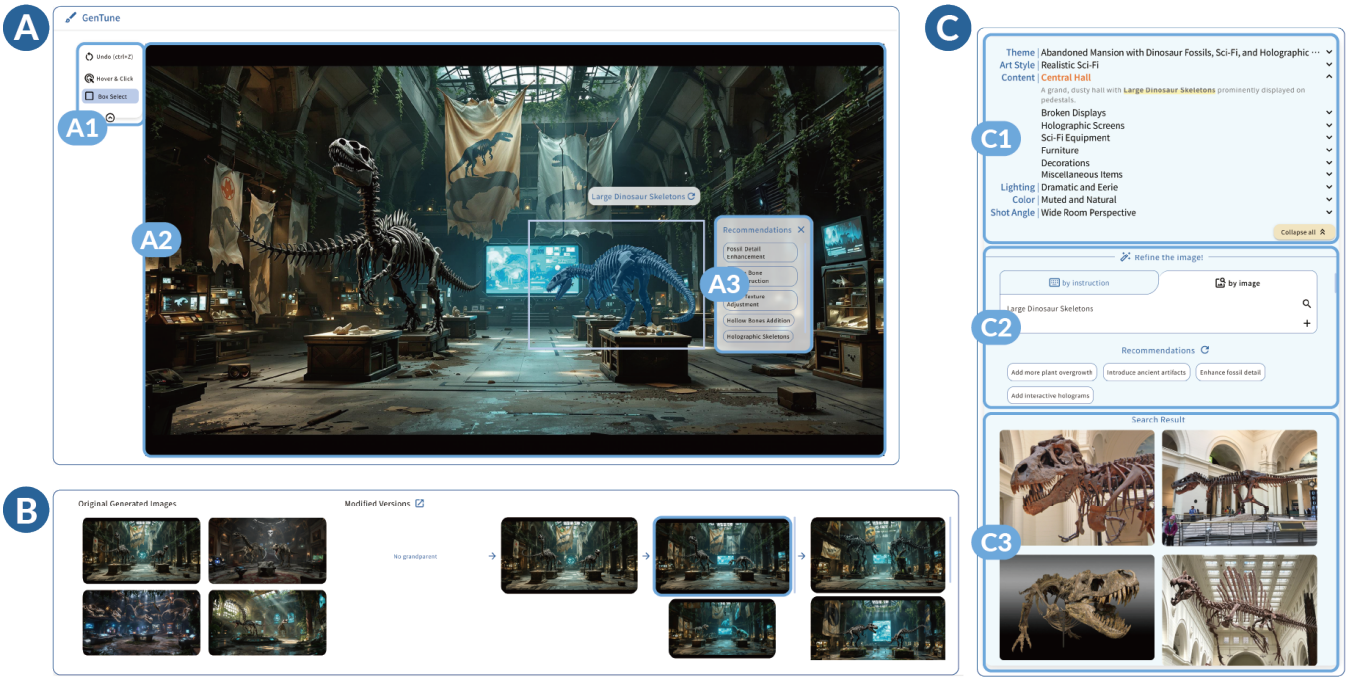

Figure 2: The main interface of GenTune includes: (A) The main image panel. Users can select elements either by hover-and-click or box-selection using the mode selector (A1). Once a region is selected, the image (A2) displays the corresponding prompt label and refinement suggestions (A3). (B) The version history view, showing the initially generated images and the iterative refinement history of the selected image. (C) The refinement panel. (C1) The structured prompt view, the corresponding section expands automatically when a region is selected. (C2) Input dialog with text input suggestions. If "by instruction" is selected, users can enter text commands to refine the image. (C3) If "by image" is selected, users can search for reference images.

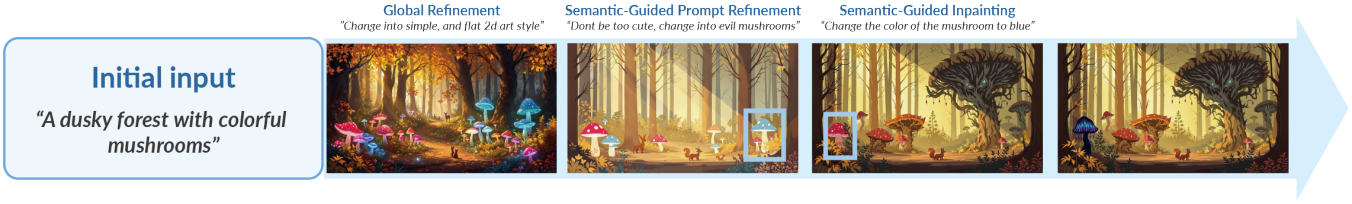

Figure 3: GenTune's refinement workflow from P14 in the open-ended study began with Global Refinement to adjust the overall style, followed by Semantic-Guided Prompt Refinement to turn all mushrooms into "evil" variants, and concluded with Semantic-Guided Inpainting for localized edits, such as changing the color of a single mushroom.

an Nvidia GeForce RTX 4080 GPU. Upon region selection, the system returned the corresponding label with an average latency of 1 second.

4.3.1 Traceable prompt. After the intial input, the system generates a structured prompt with key categories with a Brainstorming LLM (Fig. 1-A). A Label Extraction LLM extracts key elements as labels from the Content category, which then are used to define the class set for subsequent detection (Fig. 4). When a user selects an element—either by hovering and clicking or drawing a box—the system sends a point or box prompt to the SAM model [63] to obtain the corresponding segmentation mask and bounding box. These selection methods are designed to support interaction needs, allowing for either quick region selection or more precise control. To enable real-time interaction, we precompute image embeddings using the ONNX run-time version of SAM, deployed on the frontend.

Based on the bounding box, we crop the image and darken 80% of the area outside the segmentation mask. This cropped image is passed to the CLIP model [92] for semantic similarity matching. For each candidate label in the class set, we use the prompt: "The bright part is a segmentation of label" for text encoding. CLIP encodes both the image and these text descriptions into a shared embedding space, and computes similarity scores. The top five labels are returned as the predicted semantic tags for the selected region.

4.3.2 Semantic-guided refinement. For global refinement, which does not require label selection, the Global Refinement LLM processes user instructions to generate a new structured prompt.

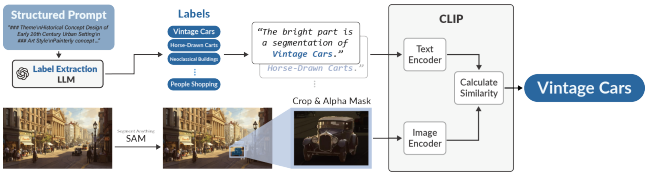

Figure 4: Prompt-element correspondence pipeline. Label Extraction LLM extracts labels from the structured prompt and used as CLIP's text inputs. When a user selects an element, it is segmented using SAM, cropped with a bounding box, alpha-masked, and passed into CLIP as the image input. CLIP then computes the text-image similarity to determine the label most associated with the selected region.

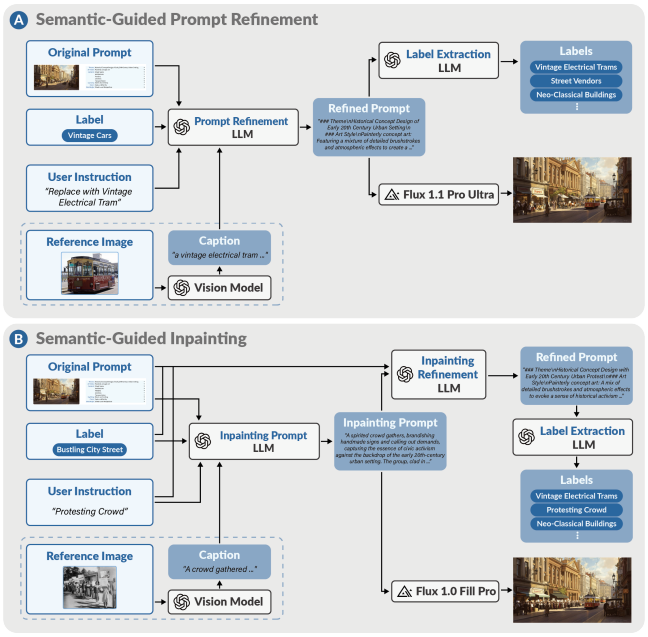

Figure 5: Technical pipeline for semantic-guided prompt refinement and inpainting. If a reference image is provided, it is captioned by a vision model. (A) The Prompt Refinement LLM takes the original prompt, extracted label, user instruction, and optional reference image caption as input to generate a refined prompt for full-scene generation. This prompt is sent to the Flux 1.1 Pro Ultra model to synthesize a new image, and concurrently is passed to the Label Extraction LLM to update semantic labels. (B) The Inpainting Prompt LLM generates a region-specific prompt using the same inputs. This prompt guides the Flux 1.0 Fill Pro model to inpaint the selected region. In parallel, the Inpainting Refinement LLM merges the new original and inpainting prompts to generate a refined prompt for label extraction.

For semantic-guided prompt refinement (Fig. 5-A), textual instructions are processed by a Refinement LLM, which takes the selected label, the user's input, and the original prompt as input. It identifies the region of the prompt associated with the label and performs highly precise modifications based on the instruction. For image-based refinement, a vision model generates a caption of the reference image, which is combined with the instruction to update the corresponding prompt segment. The refined prompt is then used to (1) generate a new image with the same seed and (2) extract updated keywords for downstream use.

For semantic-guided inpainting (Fig. 5-B), the Inpainting Prompt LLM takes similar input and generates a detailed prompt. The inpainting prompt is crafted to remain consistent with the original prompt in style and semantics. We use Flux 1.0 Fill Pro, a state-of-the-art inpainting model, to apply the modification directly to the image. Since inpainting operates directly on the image rather than modifying the prompt, it is not inherently compatible with prompt-based regeneration. To address this, GenTune reconstructs a new structured prompt from the original and inpainting prompts by the Inpainting Refinement LLM. This allows for updated label extraction and optionally enables further prompt-based generation. However, this is not the recommended workflow, as the new generation may overwrite the inpainting edits.

For refinement suggestions (Fig. 6), GenTune offers three types. Global suggestions use the original prompt to generate five variations focused on content, lighting, and atmosphere. Label-based suggestions generate three refinements and three replacements based on the selected label and its corresponding prompt segment. The expanded suggestions incorporate the user's input, the original prompt, and the selected label (if any) to produce five contextualized options.

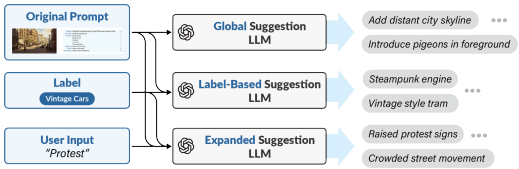

Figure 6: Pipeline for three types of suggestion LLM: Global, Label-based, and Expanded suggestions.

5 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, mirroring typical GenAI workflows. In both conditions, the full structured prompt was displayed (Fig. 2-C1). We did not

select ChatGPT as a baseline, as it is not well-suited for the complexity of environment design and is not commonly used by professionals in this domain.

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

Our study aimed to answer the following research questions:

- **RQ1:** Can GenTune help designers better interpret the relationship between prompt and image elements?
- **RQ2:** Compared to existing workflows, does GenTune enable more effective and higher-quality refinement?
- **RQ3:** Does GenTune provide greater controllability and better alignment with designers' expectation during refinement?

5.1 Study Design

5.1.1 Procedure. The study lasted approximately 2 hours and began with a 5-minute briefing. For the first task, each condition included a 10-minute tutorial, a 10-minute refinement session, and a 5-minute post-task questionnaire. System order and topics were counterbalanced across participants. The second task began with a 5-minute pre-task interview, followed by a 30-minute design session, where participants explored and applied GenTune in a self-directed manner, they completed a final questionnaire afterwards. The session concluded with a 15–20-minute post-study interview.

5.1.2 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 10 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.

5.1.3 Measurements. The post-condition questionnaire for the first task evaluated three research questions: image–prompt understanding, refinement effectiveness and quality, and overall user experience. It also included a NASA-TLX (Fig. 11) to assess perceived workload. All questions and results are shown in Figure 7. 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 [80, 86, 100, 106], and analyzed the data using the Wilcoxon signed-rank test [128], 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 8. 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 [19, 95, 109].

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 in terms of interaction flow, editing methods, refinement efficiency and quality. Participants were also asked to elaborate on their questionnaire responses through open-ended questions. All interviews were transcribed and summarized, and user interactions with GenTune were recorded. A lead author with previous experience with environment design conducted the initial coding, collaborating with two additional researchers to identify key themes. The themes were reviewed and finalized through team discussion and consensus.

5.1.4 Participants. We recruited 15 professional environment designers with 1–21 years of experience (Mean = 6.60) from industries including games (P1, P4, P6, P11–P13, P17), animation (P3, P5, P14, P21), film (P15–P16), industrial design (P20), and freelancing (P18–P19). They represented over four studios based in Japan, Singapore, and Taiwan, with clients across the EU, Japan, and the US. We also included five design students (P2, P7–P10). Nine participants (P2–P10) had taken part in the earlier formative study and were re-invited via email. The remaining participants were recruited using the same approach. All participants received $30 compensation.

6 RESULTS & FINDINGS

In this section, We report findings organized by our three research questions (RQ1–RQ3), combining results from both the controlled experiment and open-ended task.

6.1 RQ1: Prompt-to-Image Interpretability

Participants using GenTune demonstrated significantly improved prompt-to-image interpretability. In the within-subjects study, they found it significantly easier to identify the correspondence between prompts and visual elements, more effectively linked image regions to specific prompts, and better understood how each prompt influenced the image (Fig.7, Q1–Q3, $p < 0.05$). In the open-ended task, 95 and 100% of participants preferred GenTune on these aspects (Fig.8, Q1–Q3). Many (P2–8, P11–17, P19–21) noted that GenTune saved them from having to read lengthy prompts to find what to change. As P3 explained: "*The highlights clearly show what needs to be changed—I don't have to dig through the prompt.*" Others found that selecting a label clarified what would be affected (P2–6, P8–13, P16–21): "*The labels are intuitive. Once it's highlighted, I know what*

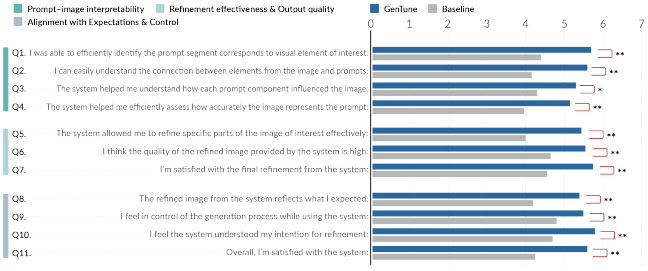

Figure 7: Survey results from the within-subject task. Participants rated Prompt-image interpretability, Refinement effectiveness and quality, and Alignment with expectations and control for both the baseline and GenTune system using a 7-point Likert scale. *: p < .05 and **: p < .01.

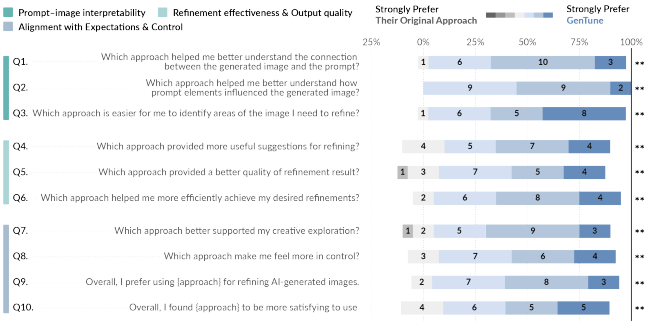

Figure 8: Survey results from the open-ended task. User rated their original approach and GenTune on a 7-point Likert scale across Prompt-image interpretability, Refinement effectiveness and quality, and Alignment with expectations and control. **: p < .01.

the system will affect. I don't have to second-guess or worry about unintended changes" (P6). P5 added, "*The labels help me understand why certain elements appear in the image.*"

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 candidate seemed plausible, but I couldn't tell which one was actually correct.*" Another noted "*Sometimes the label is inaccurate, and I have to try several times to get it right*" (P15).

6.2 RQ2: Refinement Effectiveness and Quality

6.2.1 Refinement efficiency and output quality. Participants rated GenTune significantly more effective for image refinement (Fig.7, Q5; Mean_diff = 1.2, p = 0.002), with 90% expressing a preference for it (Fig.8, Q6). 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.*"

The images refined with GenTune were significantly higher in quality and satisfaction with the final result compared to the baseline (Fig.7, Q6–Q7, p < 0.01), with 80% of participants preferring it (Fig.8, Q5). Many were impressed, as they typically relied on inpainting or manual editing (P3, P5–8, P11–14, P16–17), while GenTune's semantic-guided prompt refinement delivered better results with minimal structural disruption. As P20 noted, "*GenTune let me control the structure while making precise edits—other tools made unpredictable changes.*" However, P13 pointed out that even minor structural changes introduced by prompt refinement may be unacceptable to clients: "*Clients often require images to be 100% consistent*" Similarly, P14 preferred the aesthetic quality of their previous workflow using MidJourney.

80% of participants preferred GenTune for providing more useful refinement suggestions than their previous workflows (Fig. 8, Q4). Several noted that these suggestions introduced ideas they hadn't considered, leading to better outcomes (P3–4, P8–9, P18, P20). As P9 shared, "*When I selected the dragon, the suggestion to make its fire bigger and add a glow to the sword felt very intuitive—I loved it.*"

6.2.2 Refinement order. Since inpainting is not inherently compatible with prompt-based regeneration, just as designers rarely regenerate images after manual edits, participants often planned their strategy in advance, using prompt refinement first, followed by inpainting to avoid overwriting previous changes. As P14 explained, *"I wanted the entire forest of mushrooms to feel more sinister, then fine-tuned the color and shape of individual ones."* This approach helped maintain consistency and preserve earlier edits.

6.2.3 Within-subject case comparison. Table 2 shows the average time and number of refinement iterations for each within-subject task using the baseline and GenTune. Participants took 12.5 minutes and 6.8 iterations with the baseline, compared to 9.23 minutes and 5.3 iterations with GenTune.

Figure 10 in Appendix compares the refinement progress of P14 (GenTune) and P16 (Baseline) on topic 1. Both followed the same sequence. P16 added flowers using global refinement but relied on inpainting for the remaining edits, unsure how to precisely express the changes. He used all iterations, failed two tasks, and was dissatisfied. "*Describing edits through text wasn't intuitive. Selecting an area felt more natural—more visual thinking.*" In contrast, P14 used GenTune to select areas, assign labels, and apply mixed-mode refinement—completing all four edits in one iteration each using semantic-guided prompts refinement. "*Every step GenTune took precisely matched what I had in mind*" (P14).

Method	Avg. Time Used (min)	Avg. Iterations
Baseline	12.50	6.80
GenTune	9.23	5.30

Table 2: Average time and iterations for the within-subject task using the Baseline and GenTune methods.

6.3 RQ3: Alignment with Expectations and Control

Participants found that GenTune's refinements significantly aligned with their expectations (Fig.7, Q8; p = 0.002) "*While the output differed from my original idea, it evolved in a different direction—often exceeding my expectations*"(P9). They rated GenTune as significantly more controllable than the baseline (Fig.7, Q9; p = 0.003), with 85% preferring it (Fig.8, Q8). As P9 noted, "*Extracting the label lets me know exactly what to change. The sense of control comes from the label.*" 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.*" However, some participants noted the instablilty in GenTune during refinement "*A building I liked in the previous image disappeared after the modification*" (P21).

Participants were significantly more satisfied with GenTune compared to the baseline (Fig.7, Q11, p < 0.001) . 90% preferred GenTune for refining AI-generated images (Fig.8, Q9) 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.*" 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.*"

85% of participants preferred GenTune in supporting creative exploration (Fig.8, Q7) 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 [94].

In summary, participants found GenTune to be a more controllable and effective tool 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 design workflows.

7 FIELD STUDY

We conducted a three-day field study to evaluate how GenTune supports real-world pre-production workflows. The study focused on the efficiency and quality of the refinement process, and how GenTune helps reduce communication time between designers and stakeholders.

7.1 Participants, Study Procedure and Evaluation

We recruited three participants (Table 3): P4 (Age 27, 4 YoE) from a major game studio known for Metroidvania titles with over 3.6 million copies sold, and P22 and P3 (both Age 27, 3 YoE) from a leading visual effects studio for film and television. These were returning collaborators from our earlier studies, with upcoming projects involving new environment design tasks, and were willing to test GenTune.

We deployed the same GenTune system from the summative study via a web server, adding support for user-uploaded images during generation and refinement. During the study, participants were asked to use GenTune exclusively as their GenAI tool for pre-production tasks, including visual ideation and client communication.

Participants completed a diary study, documenting how they selected elements, issued refinement instructions, and used GenTune in their workflow. A 30-minute post-study interview followed, assessing GenTune's impact on efficiency, quality, and communication compared to their previous methods. Two researchers independently coded the interviews for thematic analysis.

ID	Field	Env. Number	Iterations
P4	Game	2	2 / 7
P22	Visual Effect	1	9
P3	Visual Effect	2	7 / 12

Table 3: Record of environments completed in the field study. In the *Iterations* column, values in the format x / y represent the number of iterations completed in the first and second environments, respectively.

7.2 Result

Table 3 shows the number of environments completed using GenTune, and total iteration counts (each iteration includes four generations or refinements).

All studios reported notable improvements in efficiency. "*I originally estimated it would take 8 hours to complete, but it only took 2.*" (P22) Two main factors contributed to this improvement. First, compared to workflows using ChatGPT and MidJourney, where one refinement often required multiple iterations, GenTune's traceable prompt structure enabled high-precision refinements. As P3 noted, "*Your tool is more direct and efficient. The local refinement feature significantly reduced the need to go back and forth with Photoshop.*" Second, GenTune accelerated communication between designers and directors. As P22 shared, "*We usually need several meetings to discuss revisions—this time, we completed three rounds of changes in half a day.*"

Figure 9 presents examples from the field study. P22 was given a two-day deadline to design a movie scene blending futuristic elements with Vietnamese rice fields with the output used directly for 3D scene setup. Starting with the prompt "*Design a Vietnamese farming village at sunset, with rice fields and high-tech devices*", the client selected one of the generated images and requested two revisions: (1) replace futuraistic buildings with traditional architecture, and (2) remove the person and add a giant armored robot. Using GenTune, the designer completed both changes in 9 iterations, and the final direction was approved. P4 was tasked with designing a ceiling and room layout based on a reference image, with the output serving as a guide for 3D environment development. For the ceiling, she selected the window area and, in just two iterations, achieved a structure suitable for further concept design.

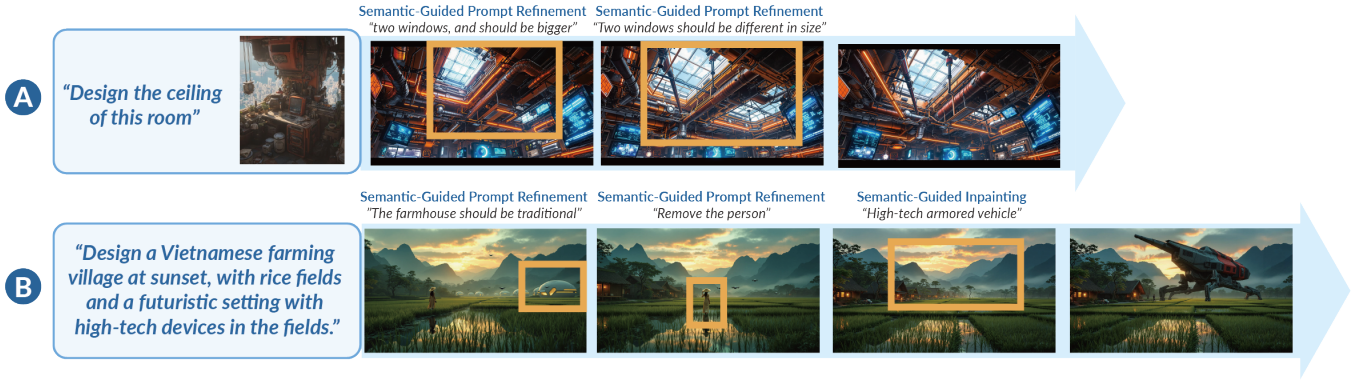

Figure 9: Workflow procedure and results of using GenTune on real-world projects from two designers (P4 and P22). (A) P4 designed a ceiling and room layout based on a reference image using semantic-guided prompt refinement. (B) P22 designed a movie scene blending futuristic elements with Vietnamese rice fields using a combination of semantic-guided prompt refinement and inpainting.

Participants also praised GenTune for enhancing both quality and creativity. As P4 noted, "*I originally thought two skylights were enough, but GenTune suggested making them asymmetrical—and it worked surprisingly well.*" P3 added, "*It often brings out colors or moods I hadn't considered, which helps give the director an early visual impression.*" Some participants also described unexpected but efficient use cases, "*I can quickly generate props and textures with GenTune and bash them into my draft*" (P3). However, some challenges remained. P22 shared, "*I generated over twenty images to get the giant mech I wanted because the client had already approved the layout. Adding the mech through prompts caused too much structural change.*" Still, P22 acknowledged that GenTune significantly reduced overall revision and communication time.

We are excited to share that following the field study, both studios have continued to use GenTune in their commercial projects, producing over 40 environments to date. This demonstrates GenTune's strong potential for real-world adoption and seamless integration into professional creative workflows.

8 DISCUSSION, LIMITATIONS, AND FUTURE WORK

8.1 Adapting Generation Approaches to Designers' Needs

In the within-subject study, among 80 total refinement actions by 20 participants, 47 involved prompt refining, while 33 used inpainting. They attributed their preference to improvements in quality and consistency. Figure 13 in Appendix A shows comparisons from our open-ended study, where each image pair was generated with inpainting and with prompt refinement. "*When adding people, the lighting and posture look more consistent and natural with the seed*" (Appendix A, Fig.13-A, 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*"(Appendix A, Fig. 13-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"[7], 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 (Appendix A, Fig.14-A), but semantic-guided refinement replaced all instances associated with the "blackboard" label. Similarly, when the prompt modification is broad, fine details may be lost. In Appendix A, Figure 14-B, P12 used the Izumo Taisha shrine as input reference and wanted to add the traditional shimenawa (straw rope). In the prompt refinement result, while the shrine's structure was preserved, the torii gate disappeared. In contrast, inpainting resulting in an outcome that better met the designer's expectations.

We intentionally omitted spatial conditioning controls during refinement, while potentially offering precise structural control, they risk issues such as detail loss, texture degradation, and unintended style shifts. These inconsistencies can disrupt the overall aesthetic coherence, which is critical for designers. Figure 15 in Appendix A compares the results of the GenTune, Flux 1.0 (Depth) and the ChatGPT image editing feature (released 3/25), using the same input: "add prayer flags to the towers". Both ChatGPT and Flux outputs show significant architectural detail loss, drastic changes in style and texture, and even the disappearance of key elements like the bridge.

This design implication extends to other creative domains, where different disciplines prioritize different needs and may benefit from distinct generative models. For example, interior designers often require high structural accuracy [121], making spatial conditioning controls essential. In contrast, graphic designers may prioritize layout composition [27], suggesting that region-based spatial control models may be more appropriate.

8.2 Generalizing GenTune-Traceable Control in an Era of Automated Generation

GenTune's core concept is a model-agnostic HCI paradigm that enables traceable, element-level control in human–AI collaboration. Our work addresses a critical challenge emerging from state-of-the-art generative workflows, where the automated translation of high-level concepts into complex intermediate representations before producing a final output can limit a user's ability to understand, steer, and refine the generative process [25]. This is especially true in fast-paced settings like environment design, where designers must rapidly produce variations, making the manual rewriting of prompts impractical.

This challenge applies broadly, from image generation [18, 122] to video, where LLMs act as directors or motion planners [51, 82, 111], and programming, where LLMs handle planning [59], multi-step guidance [49], or specification generation [48] prior to code synthesis. As these workflows become more automated and multi-step, particularly with the rise of multi-agent systems and MLLMs[24, 33, 99, 142], the need for understandability and controllability of intermediate outputs becomes increasingly critical.

Our paradigm introduces traceable links between intermediate representations and final results, enabling users to interpret and revise intermediate steps directly from the output. This traceability significantly enhances designers' sense of control, transparency, trust, and alignment with creative intent in this work. supporting core principles of human-centered AI [8, 105, 132].

GenTune's implementation can also be directly applied to other complex visual domains that require fine-grained control. For example, P14 successfully used GenTune for character design, effectively controlling individual elements. Similar potential exists in interior design, where elements such as walls, floors, and furniture can be treated independently [17], and in game UI design, where interface components are inherently structured and easily labeled [14]. Another promising direction is video and animation generation, where evolving visual elements can be labeled with attributes like motion type, speed, and direction. These can be independently adjusted, aligning with established practices in motion design and layered animation [119], and opening opportunities for traceable control in temporal media.

While our current implementation supports caption-based image inputs, future work could extend the traceable concept to sketches and reference images, enabling users to map generated elements back to specific visual regions, offering even more granular control.

8.3 Ethical Concern

As GenAI becomes increasingly embedded in creative workflows [65], prior research has raised concerns about its negative impacts, such as the displacement of professional artists [58, 116] and the growing pressure for designers to adopt AI tools to stay competitive [104, 116].

Most of our participants have formal training in digital art and are already expected to integrate GenAI tools into their professional pipelines. GenTune was developed in direct response to the needs and challenges voiced by working designers, and is designed for referencing and facilitating communication, not as a final asset. We see designers as the core creative force, and GenTune aims to augment their workflows by improving control and reducing trial-and-error, allowing them to focus on creative decisions that require their expertise.

At the same time, we recognize that even positive efficiency gains can contribute to increased client demands, potentially reinforcing the "treadmill" effect. Understanding how AI tools reshape client–designer dynamics is a critical direction for future ethnographic research.

Concerns have also been raised about GenAI reducing group creativity [34, 66]. In practice, environment designers "*actively conduct research and curate references, filtering GenAI outputs to integrate their own insights*", stated by a participant with 13 years of experience (P18). In this sense, systems like GenTune can amplify creative exploration and foster greater diversity through collaboration [45]. However, we acknowledge that overreliance on GenAI may risk homogenization, an important issue for future long-term research on human–AI co-creation.

8.4 Limitations and Future Work

Study design. We evaluated GenTune through an open-ended task in which participants applied it to a current or past project, relying primarily on self-reported data. While this approach offers ecological validity, variations in participants' workflows and prior GenAI experience may introduce inconsistencies when comparing GenTune to their original methods. To more rigorously assess refinement effectiveness and output quality, future studies could incorporate a standardized baseline and involve external experts to evaluate and compare refinement results, enabling more insightful analysis.

Refinement order. While combining inpainting and prompt refinement offers flexibility, it 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. As models advance, integrating spatial conditioning that merges prompt refinement results with inpainted regions could offer a more seamless solution.

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 [11, 21]. Recent work has begun addressing these issues through improved prompt refinement [83], enhanced spatial understanding [21], and multi-view consistency [50]. The development of spatially conditioned control models [143] 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.

9 CONCLUSION

We present GenTune, a human-centered generative AI system and model-agnostic HCI paradigm that enables traceable, element-level control to improve the understandability and controllability of human–AI collaboration. Designed for environment design workflows, GenTune combines traceable prompts and semantic-guided refinement to help designers better interpret prompt–image relationships and perform more precise, consistent edits. We evaluated GenTune through a summative study with 20 environment designers, including a within-subject experiment and an open-ended design task. Results showed that GenTune significantly improved prompt–image comprehension, refinement quality, efficiency, and user control—receiving strong preference over existing workflows. A follow-up field study in two professional studios further demonstrated GenTune's potential to enhance refinement efficiency and creative communication in real-world production settings.

Acknowledgments

This work was supported by the National Science and Technology Council, Taiwan (NTSC 112-2221-E-002-185-MY3) and the Center of Data Intelligence: Technologies, Applications, and Systems at National Taiwan University (113L900901, 113L900902, 113L900903), funded through the Featured Areas Research Center Program under the Higher Education Sprout Project by the Ministry of Education (MOE) of Taiwan. We also acknowledge support from National Taiwan University, Moonshine Studio, Winking Studios, and Rayark Games. Finally, we extend our gratitude to all participants and reviewers for their valuable feedback.

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] Krzysztof Adamkiewicz, Paweł Wojciech Woźniak, Julia Dominiak, Andrzej Romanowski, Jakob Karolus, and Stanislav Frolov. 2025. PromptMap: An Alternative Interaction Style for AI-Based Image Generation. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25).* Association for Computing Machinery, New York, NY, USA, 1162–1176. https://doi.org/10.1145/3708359.3712150

[4] Robin S Adams and Cynthia J Atman. 1999. Cognitive processes in iterative design behavior. In *FIE'99 Frontiers in Education. 29th Annual Frontiers in Education Conference. Designing the Future of Science and Engineering Education. Conference Proceedings (IEEE Cat. No. 99CH37011*, Vol. 1. IEEE, 11A6–13.

[5] Noor Wahyuni Ahmad and Suzana Ruslan. 2024. Crafting Effective Prompts: A Guideline for Successful Image Generation. In *2024 14th International Conference on System Engineering and Technology (ICSET).* IEEE, 84–89.

[6] Shm Garanganao Almeda, JD Zamfirescu-Pereira, Kyu Won Kim, Pradeep Mani Rathnam, and Bjoern Hartmann. 2024. Prompting for Discovery: Flexible Sense-Making for AI Art-Making with Dreamsheets. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–17.

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

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

[9] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21).* Association for Computing Machinery, New York, NY, USA, Article 81, 16 pages. https://doi.org/10.1145/3411764.3445717

[10] 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.

[11] 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.

[12] Fouad Bousetouane. 2025. Generative AI for Vision: A Comprehensive Study of Frameworks and Applications. *arXiv preprint arXiv:2501.18033* (2025).

[13] 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.

[14] Simon Brewer. 2023. *The Pocket Mentor for Video Game UX UI (The Pocket Mentors for Games Careers)* (1st ed.).

[15] 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.

[16] Tim Brooks, Aleksander Holynski, and Alexei A Efros. 2023. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 18392–18402.

[17] L-S Byun. 2006. Peculiarity of interior design materials for accommodation areas of cruise ships: A state-of-the-art review. *Ships and Offshore Structures* 1, 3 (2006), 171–183.

[18] 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.

[19] 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.

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

[21] 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.

[22] Chen Chen, Cuong Nguyen, Thibault Groueix, Vladimir G. Kim, and Nadir Weibel. 2024. MemoVis: A GenAI-Powered Tool for Creating Companion Reference Images for 3D Design Feedback. *ACM Trans. Comput.-Hum. Interact.* 31, 5, Article 67 (Nov. 2024), 41 pages. https://doi.org/10.1145/3694681

[23] 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.* 1–19.

[24] 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).

[25] 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).

[26] YU-HAN CHIU and Chun-Ching Chen. 2024. Elevating and Sharpening Convergent Thinking: The Potential of Generative AI for Creative Professionals. *Available at SSRN 4911494* (2024).

[27] 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.* 1–25.

[28] Leah Chong, I-Ping Lo, Jude Rayan, Steven Dow, Faez Ahmed, and Ioanna Lykourentzou. 2025. Prompting for products: investigating design space exploration strategies for text-to-image generative models. *Design Science* 11 (2025), e2.

[29] 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* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 6, 17 pages. https://doi.org/10.1145/3586183.3606777

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

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

[32] Sven Coppers, Jan Van den Bergh, Kris Luyten, Karin Coninx, Iulianna van der Lek-Ciudin, Tom Vanallemeersch, and Vincent Vandeghinste. 2018. Intellingo: An Intelligible Translation Environment. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174098

[33] 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).

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

[35] 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.

[36] Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, and Rajiv Ranjan. 2023. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Comput. Surv.* 55, 9, Article 194 (Jan. 2023), 33 pages. https://doi.org/10.1145/3561048

[37] Upol Ehsan, Philipp Wintersberger, Q. Vera Liao, Elizabeth Anne Watkins, Carina Manger, Hal Daumé III, Andreas Riener, and Mark O Riedl. 2022. Human-Centered Explainable AI (HCXAI): Beyond Opening the Black-Box of AI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 109, 7 pages. https://doi.org/10.1145/3491101.3503727

[38] 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).

[39] Noyan Evirgen and Xiang'Anthony' Chen. 2022. Ganzilla: User-driven direction discovery in generative adversarial networks. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–10.

[40] Noyan Evirgen and Xiang'Anthony Chen. 2023. Ganravel: User-driven direction disentanglement in generative adversarial networks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[41] Noyan Evirgen, Ruolin Wang, and Xiang 'Anthony Chen. 2024. From Text to Pixels: Enhancing User Understanding through Text-to-Image Model Explanations. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) *(IUI '24)*. Association for Computing Machinery, New York, NY, USA, 74–87. https://doi.org/10.1145/3640543.3645173

[42] Yingchaojie Feng, Xingbo Wang, Kam Kwai Wong, Sijia Wang, Yuhong Lu, Minfeng Zhu, Baicheng Wang, and Wei Chen. 2023. Promptmagician: Interactive prompt engineering for text-to-image creation. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2023), 295–305.

[43] Jonas Frich, Midas Nouwens, Kim Halskov, and Peter Dalsgaard. 2021. How Digital Tools Impact Convergent and Divergent Thinking in Design Ideation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 431, 11 pages. https://doi.org/10.1145/3411764.3445062

[44] Jiaying Fu, Xiruo Wang, Zhouyi Li, Kate Vi, Chuyan Xu, and Yuqian Sun. 2025. " I Like Your Story!": A Co-Creative Story-Crafting Game with a Persona-Driven Character Based on Generative AI. *arXiv preprint arXiv:2503.09102* (2025).

[45] Manuel B Garcia. 2024. The paradox of artificial creativity: Challenges and opportunities of generative AI artistry. *Creativity Research Journal* (2024), 1–14.

[46] 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).

[47] Yuhan Guo, Hanning Shao, Can Liu, Kai Xu, and Xiaoru Yuan. 2024. Promptthis: Visualizing the process and influence of prompt editing during text-to-image creation. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[48] Hojae Han, Jaejin Kim, Jaeseok Yoo, Youngwon Lee, and Seung-won Hwang. 2024. Archcode: Incorporating software requirements in code generation with large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 13520–13552.

[49] Yewei Han and Chen Lyu. 2025. Multi-stage guided code generation for Large Language Models. *Engineering Applications of Artificial Intelligence* 139 (2025), 109491.

[50] 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.

[51] Susung Hong, Junyoung Seo, Heeseong Shin, Sunghwan Hong, and Seungryong Kim. 2023. Direct2v: Large language models are frame-level directors for zero-shot text-to-video generation. *arXiv preprint arXiv:2305.14330* (2023).

[52] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. 2021. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations*. https://arxiv.org/abs/2106.09685

[53] 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).

[54] Rong Huang, Haichuan Lin, Chuanzhang Chen, Kang Zhang, and Wei Zeng. 2024. PlantoGraphy: Incorporating Iterative Design Process into Generative Artificial Intelligence for Landscape Rendering. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 168, 19 pages. https://doi.org/10.1145/3613904.3642824

[55] Mina Huh, Dingzeyu Li, Kim Pimmel, Hijung Valentina Shin, Amy Pavel, and Mira Dontcheva. 2025. VideoDiff: Human-AI Video Co-Creation with Alternatives. *arXiv preprint arXiv:2502.10190* (2025).

[56] Syed Fahad Javaid and James Paul Pandarakalam. 2021. The association of creativity with divergent and convergent thinking. *Psychiatria danubina* 33, 2 (2021), 133–139.

[57] Youngseung Jeon, Seungwan Jin, Patrick C. Shih, and Kyungsik Han. 2021. FashionQ: An AI-Driven Creativity Support Tool for Facilitating Ideation in Fashion Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 576, 18 pages. https://doi.org/10.1145/3411764.3445093

[58] 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* (Montréal, QC, Canada) *(AIES '23)*. Association for Computing Machinery, New York, NY, USA, 363–374. https://doi.org/10.1145/3600211.3604681

[59] Xue Jiang, Yihong Dong, Lecheng Wang, Zheng Fang, Qiwei Shang, Ge Li, Zhi Jin, and Wenpin Jiao. 2024. Self-planning code generation with large language models. *ACM Transactions on Software Engineering and Methodology* 33, 7 (2024), 1–30.

[60] Eric William Johnson. 1997. *Analysis and refinement of iterative design processes*. University of Notre Dame.

[61] Kyung Hee Kim and Robert A Pierce. 2013. Convergent versus divergent thinking. *Encyclopedia of creativity, invention, innovation and entrepreneurship* (2013), 245–250.

[62] Sunnie S. Y. Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. "Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 250, 17 pages. https://doi.org/10.1145/3544548.3581001

[63] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4015–4026.

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

[65] Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. 2023. Understanding Visual Artists' Adoption of Large-scale Text-to-image Generation Models for Creative Works. In *Proceedings of the 2023 ACM SIGCHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3581641.3584078

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

[67] Shuangqi Li, Hieu Le, Jingyi Xu, and Mathieu Salzmann. 2024. Enhancing Compositional Text-to-Image Generation with Reliable Random Seeds. *arXiv*

preprint arXiv:2411.18810 (2024).

[68] Xian Li, Yuanning Han, Di Liu, Pengcheng An, and Shuo Niu. 2024. FlowGPT: Exploring Domains, Output Modalities, and Goals of Community-Generated AI Chatbots. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*. 355–361.

[69] Q Vera Liao and Kush R Varshney. 2021. Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790* (2021).

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

[71] David Chuan-En Lin, Hyeonsu B Kang, Nikolas Martelaro, Aniket Kittur, Yan-Ying Chen, and Matthew K Hong. 2025. Inkspire: Supporting Design Exploration with Generative AI through Analogical Sketching. *arXiv preprint arXiv:2501.18588* (2025).

[72] 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).

[73] Joseph Lindley and Roger Whitham. 2025. From Prompt Engineering to Prompt Craft. In *Proceedings of the Nineteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–12.

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

[75] 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* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 73, 17 pages. https://doi.org/10.1145/3526113.3545621

[76] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating Text-to-Image AI in 3D Design Workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1955–1977. https://doi.org/10.1145/3563657.3596098

[77] 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.

[78] 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.

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

[80] 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.

[81] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. 2022. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 11461–11471.

[82] Jiaxi Lv, Yi Huang, Mingfu Yan, Jiancheng Huang, Jianzhuang Liu, Yifan Liu, Yafei Wen, Xiaoxin Chen, and Shifeng Chen. 2024. Gpt4motion: Scripting physical motions in text-to-video generation via blender-oriented gpt planning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1430–1440.

[83] 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).

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

[85] Jonas Oppenlaender, Rhema Linder, and Johanna Silvennoinen. 2024. Prompting AI art: An investigation into the creative skill of prompt engineering. *International journal of human–computer interaction* (2024), 1–23.

[86] 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.

[87] Srishti Palani and Gonzalo Ramos. 2024. Evolving Roles and Workflows of Creative Practitioners in the Age of Generative AI. In *Proceedings of the 16th Conference on Creativity & Cognition* (Chicago, IL, USA) *(C&C '24)*. Association for Computing Machinery, New York, NY, USA, 170–184. https://doi.org/10.1145/3635636.3656190

[88] Hyerim Park, Malin Eiband, Andre Luckow, and Michael Sedlmair. 2025. Exploring Visual Prompts: Refining Images with Scribbles and Annotations in Generative AI Image Tools. *arXiv preprint arXiv:2503.03398* (2025).

[89] 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

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

[91] 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.

[92] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PmLR, 8748–8763.

[93] Muhammad Raees, Inge Meijerink, Ioanna Lykourentzou, Vassilis-Javed Khan, and Konstantinos Papangelis. 2024. From explainable to interactive AI: A literature review on current trends in human-AI interaction. *International Journal of Human-Computer Studies* (2024), 103301.

[94] 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).

[95] 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.

[96] Ana Rodrigues, Diogo Cabral, and Pedro F. Campos. 2023. Creativity Support Tools and Convergent Thinking: A Preliminary Review on Idea Evaluation and Selection. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) *(C&C '23)*. Association for Computing Machinery, New York, NY, USA, 305–311. https://doi.org/10.1145/3591196.3596821

[97] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[98] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2023. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 22500–22510.

[99] 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).

[100] 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.

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

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

[103] Pol Baladas Gerard Serra, Oriol Domingo, and Pol Baladas. 2022. A programmable interface for creative exploration. In *Machine Learning for Creativity and Design Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022)(December 2022). https://neuripscreativityworkshop. github. io/2022/papers/ml4cd2022_paper19. pdf.*

[104] 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).

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

[106] 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.

[107] Ileana Stigliani and Davide Ravasi. 2018. The shaping of form: Exploring designers' use of aesthetic knowledge. *Organization Studies* 39, 5-6 (2018), 747–784.

[108] Yuan Sun, Eunchae Jang, Fenglong Ma, and Ting Wang. 2024. Generative AI in the Wild: Prospects, Challenges, and Strategies. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 747, 16 pages. https://doi.org/10.1145/3613904.3642160

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

[110] Linus Tan and Max Luhrs. 2024. Using Generative AI Midjourney to enhance divergent and convergent thinking in an architect's creative design process. *The Design Journal* 27, 4 (2024), 677–699.

[111] Shuai Tan, Biao Gong, Yutong Feng, Kecheng Zheng, Dandan Zheng, Shuwei Shi, Yujun Shen, Jingdong Chen, and Ming Yang. 2025. Mimir: Improving video diffusion models for precise text understanding. In *Proceedings of the Computer*

Vision and Pattern Recognition Conference. 23978–23988.

[112] 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).

[113] Anja Thieme, Ed Cutrell, Cecily Morrison, Alex Taylor, and Abigail Sellen. 2020. Interpretability as a dynamic of human-AI interaction. *Interactions* 27, 5 (Sept. 2020), 40–45. https://doi.org/10.1145/3411286

[114] Tiffany Tseng, Ruijia Cheng, and Jeffrey Nichols. 2024. Keyframer: Empowering animation design using large language models. *arXiv preprint arXiv:2402.06071* (2024).

[115] Usman Ahmad Usmani, A. Happonen, and J. Watada. 2023. Human-Centered Artificial Intelligence: Designing for User Empowerment and Ethical Considerations. *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (2023), 01–05. https://doi.org/10.1109/HORA58378.2023.10156761

[116] Veera Vimpari, Annakaisa Kultima, Perttu Hämäläinen, and Christian Guckelsberger. 2023. "An Adapt-or-Die Type of Situation": Perception, Adoption, and Use of Text-to-Image-Generation AI by Game Industry Professionals. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY (2023), 131–164.

[117] Zijun Wan, Jiawei Tang, Linghang Cai, Xin Tong, and Can Liu. 2024. Breaking the Midas Spell: Understanding Progressive Novice-AI Collaboration in Spatial Design. *arXiv preprint arXiv:2410.20124* (2024).

[118] Fengxiang Wang, Wanrong Huang, Shaowu Yang, Qi Fan, and Long Lan. 2024. Learning to learn better visual prompts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 5354–5363.

[119] John YA Wang and Edward H Adelson. 1993. Layered representation for motion analysis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 361–366.

[120] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B Chilton. 2023. PopBlends: Strategies for Conceptual Blending with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 435, 19 pages. https://doi.org/10.1145/3544548.3580948

[121] Shun-Yu Wang, Wei-Chung Su, Serena Chen, Ching-Yi Tsai, Marta Misztal, Katherine M Cheng, Alwena Lin, Yu Chen, and Mike Y Chen. 2024. RoomDreaming: Generative-AI Approach to Facilitating Iterative, Preliminary Interior Design Exploration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.

[122] 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).

[123] Xinru Wang and Ming Yin. 2021. Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making. In *Proceedings of the 26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21)*. Association for Computing Machinery, New York, NY, USA, 318–328. https://doi.org/10.1145/3397481.3450650

[124] 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.

[125] Yunlong Wang, Priyadarshini Venkatesh, and Brian Y Lim. 2022. Interpretable Directed Diversity: Leveraging Model Explanations for Iterative Crowd Ideation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 183, 28 pages. https://doi.org/10.1145/3491102.3517551

[126] 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

[127] Jingxuan Wei, Shiyu Wu, Xin Jiang, and Yequan Wang. 2023. Dialogpaint: A dialog-based image editing model. *arXiv preprint arXiv:2303.10073* (2023).

[128] Robert F Woolson. 2005. Wilcoxon signed-rank test. *Encyclopedia of biostatistics* 8 (2005).

[129] Hsuan-Yi Wu and Vic Callaghan. 2016. From imagination to innovation: a creative development process. In *Intelligent Environments 2016*. IOS Press, 514–523.

[130] X Xie. 2023. The cognitive process of creative design: a perspective of divergent thinking. Think Skills Creat. 101266 (2023).

[131] Katherine Xu, Lingzhi Zhang, and Jianbo Shi. 2024. Good seed makes a good crop: Discovering secret seeds in text-to-image diffusion models. *arXiv preprint arXiv:2405.14828* (2024).

[132] 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.

[133] Xiangyuan Xue, Zeyu Lu, Di Huang, Wanli Ouyang, and Lei Bai. 2024. GenAgent: Build Collaborative AI Systems with Automated Workflow Generation–Case Studies on ComfyUI. *arXiv preprint arXiv:2409.01392* (2024).

[134] Zihan Yan, Chunxu Yang, Qihao Liang, and Xiang'Anthony' Chen. 2023. XCreation: A Graph-based Crossmodal Generative Creativity Support Tool. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[135] 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.

[136] Tao Yu, Runseng Feng, Ruoyu Feng, Jinming Liu, Xin Jin, Wenjun Zeng, and Zhibo Chen. 2023. Inpaint anything: Segment anything meets image inpainting. *arXiv preprint arXiv:2304.06790* (2023).

[137] J.D. Zamfirescu-Pereira, Heather Wei, Amy Xiao, Kitty Gu, Grace Jung, Matthew G Lee, Bjoern Hartmann, and Qian Yang. 2023. Herding AI Cats: Lessons from Designing a Chatbot by Prompting GPT-3. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2206–2220. https://doi.org/10.1145/3563657.3596138

[138] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. https://doi.org/10.1145/3544548.3581388

[139] Chengzhi Zhang, Weijie Wang, Paul Pangaro, Nikolas Martelaro, and Daragh Byrne. 2023. Generative image AI using design sketches as input: Opportunities and challenges. In *Proceedings of the 15th Conference on Creativity and Cognition*. 254–261.

[140] Fang Zhang, Zhenlun Sun, and Qian Chen. 2024. Research on Interior Intelligent Design System Based On Image Generation Technology. *Procedia Computer Science* 243 (2024), 690–699.

[141] 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*. 1–18.

[142] Jiayi Zhang, Jinyu Xiang, Zhaoyang Yu, Fengwei Teng, Xionghui Chen, Jiaqi Chen, Mingchen Zhuge, Xin Cheng, Sirui Hong, Jinlin Wang, et al. 2024. Aflow: Automating agentic workflow generation. *arXiv preprint arXiv:2410.10762* (2024).

[143] 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.

[144] Weili Zhu, Siyuan Shang, Weili Jiang, Meng Pei, and Yanjie Su. 2019. Convergent thinking moderates the relationship between divergent thinking and scientific creativity. *Creativity Research Journal* 31, 3 (2019), 320–328.

A Appendix

Figure 10: Comparison of refinement workflows between the Baseline and GenTune systems for one of the topics: The Hanging Gardens of Neo-Babylon in the within-subject task. Participants were shown an initial generated image (left column) and asked to perform four refinement tasks. The center column illustrates the Baseline workflow, which relied on prompt modification and inpainting. The right column shows GenTune's workflow and the result of user-chosen refinement method between semantic-guided prompt refinement and inpainting.

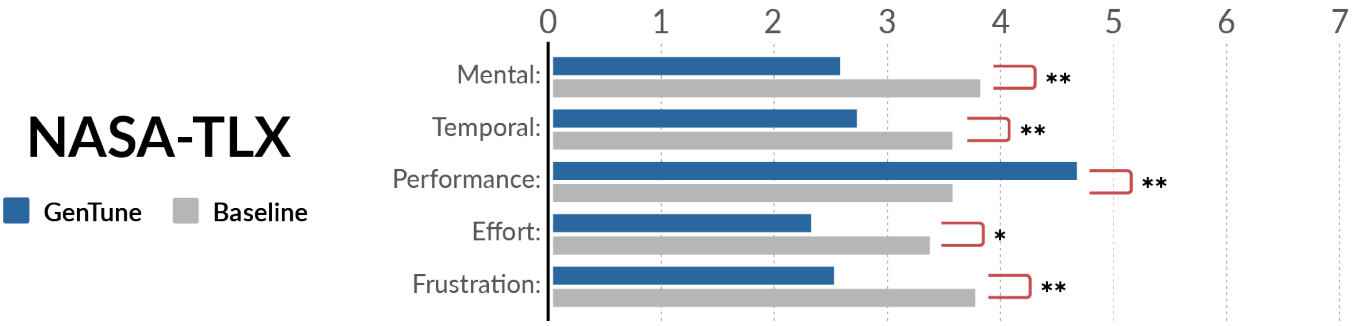

Figure 11: Survey results from the within-subject task. Participants rated the NASA-TLX workload for both the baseline and GenTune system using a 7-point Likert scale. *: p < .05 and **: p < .01.

Figure 12: Results from the open-ended task. Each row presents iterative image refinements made by different users using GenTune, including semantic-guided prompt refinement via text instruction or image reference, as well as semantic-guided inpainting.

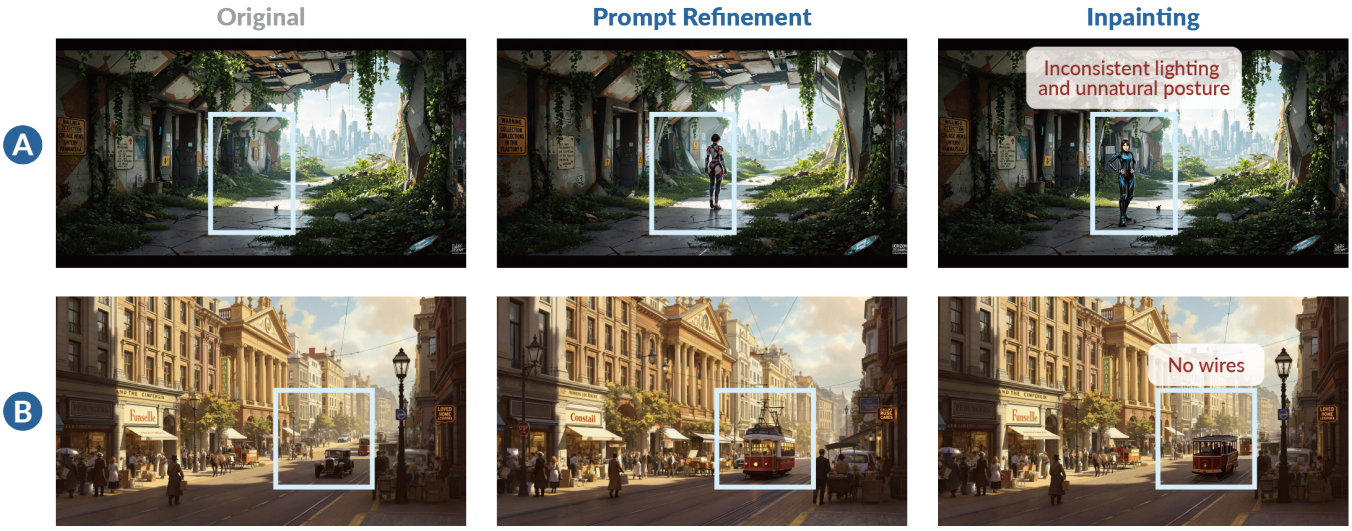

Figure 13: Comparison between semantic-guided prompt refinement with controlled seed (middle) and semantic-guided inpainting (right). In these examples, seed-based refinement better preserves overall image aesthetic coherence.

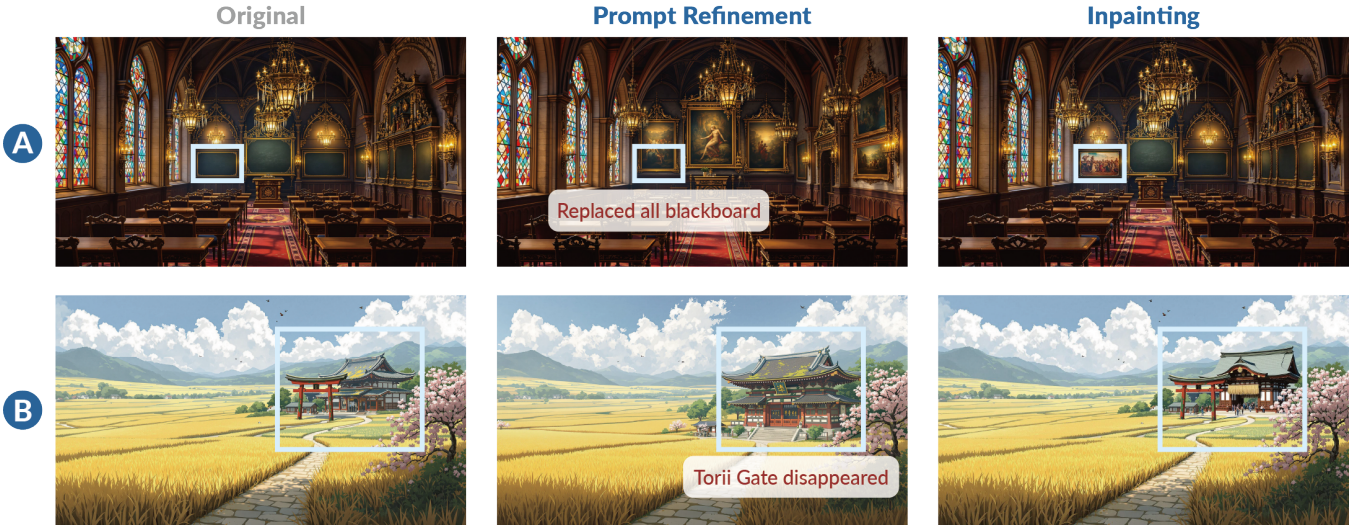

Figure 14: Comparison between semantic-guided prompt refinement with controlled seed (middle) and semantic-guided inpainting (right). In this case, inpainting method provides more precise control.

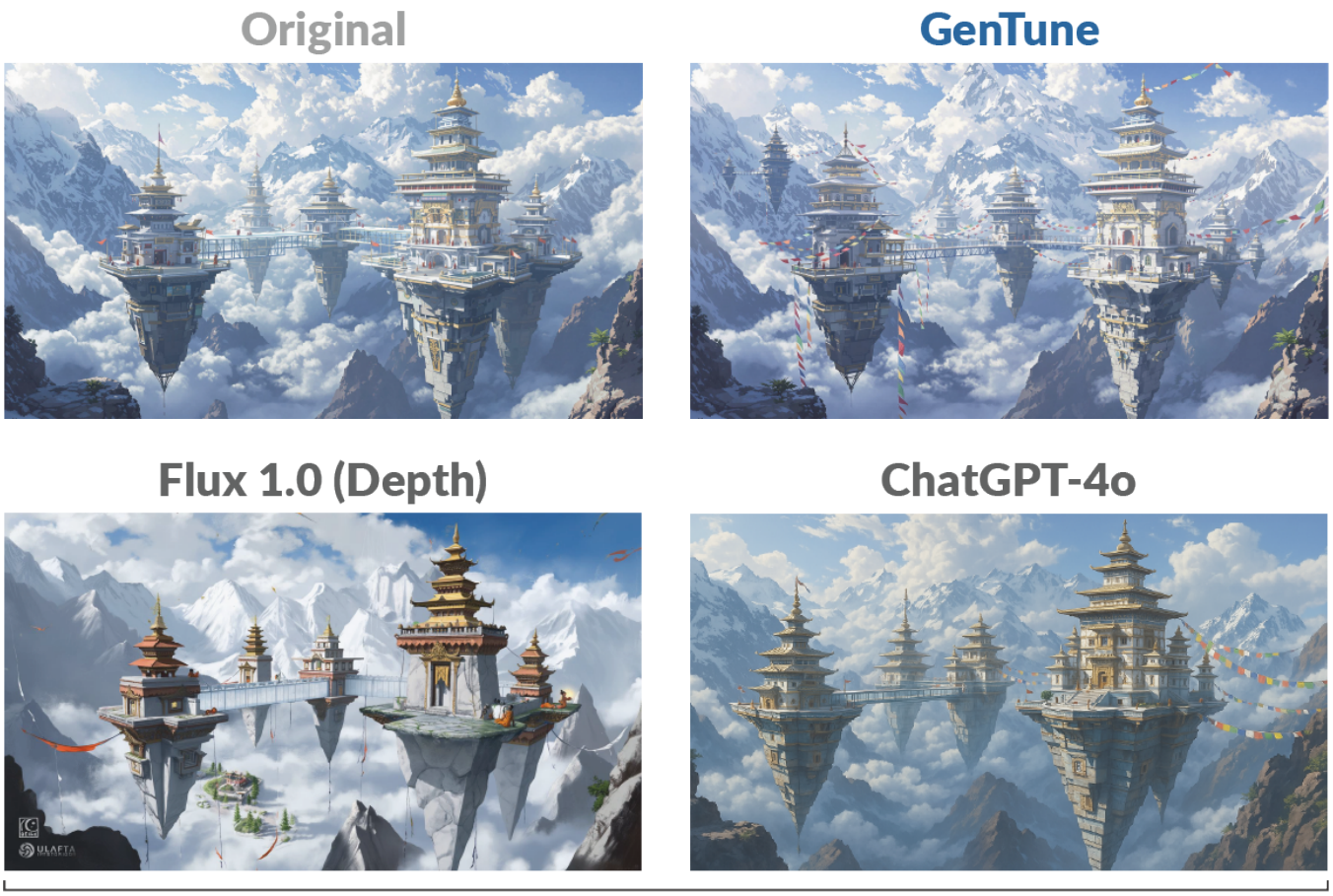

Figure 15: Comparison of image refinement results among GenTune, Flux 1.0 (Depth), and ChatGPT-4o. All methods were given the same instruction: "Add prayer flags on the towers." GenTune preserves visual coherence, while Flux and ChatGPT exhibit detail loss, drastic changes in style and texture, and even the disappearance of key elements (e.g., the bridge).