Interaction Methods in Generative AI Image Tools: A Review of Trends and Design Opportunities Across HCI and Industry

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

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

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

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

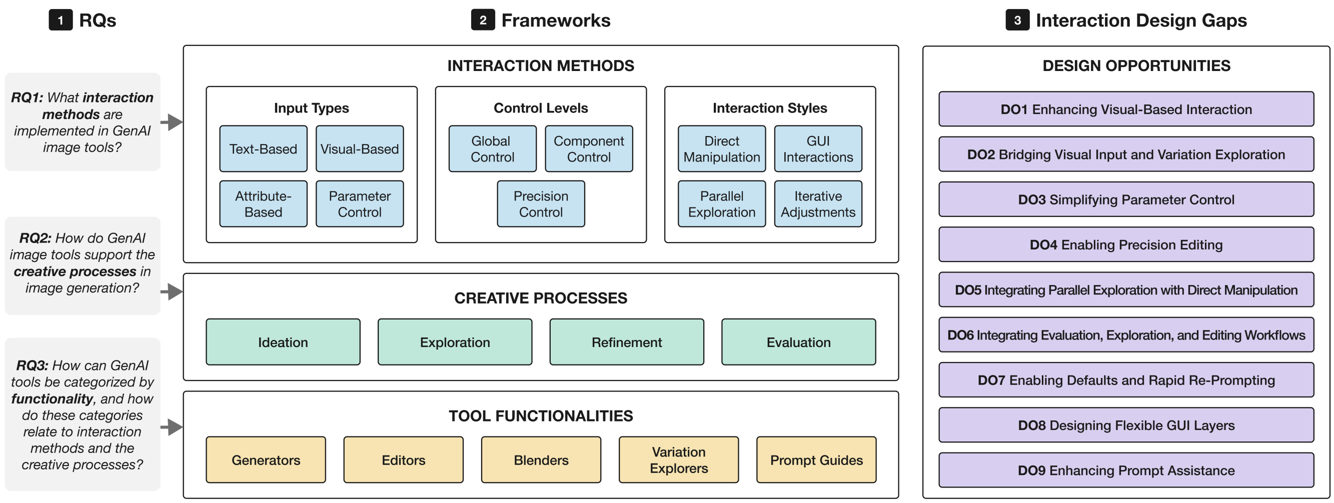


Figure 1: Overview of our design space for GenAI image tool interfaces. The figure illustrates (1) our three research questions (RQ1–RQ3), (2) the analytical framework used to address them—consisting of *interaction methods, creative processes,* and *tool functionalities*—and (3) the nine design opportunities (DO1–DO9) that emerge from gaps identified across these dimensions.

Abstract

Generative AI (GenAI) image tools are increasingly integrated into design workflows, prompting HCI research on their interaction methods and interfaces. We reviewed 37 such tools, including 28 HCI research systems and nine commercial systems (2022–July 2025), using three analytical frameworks: interaction methods, creative processes, and tool functionalities. We found that text prompts remain the dominant input method, while visual and attribute-based inputs—particularly in academic tools—are gaining traction and are often combined with text for refinement. Commercial systems emphasize parameter control, whereas academic tools focus on semantic attributes and visual organization. Most tools support ideation and exploration, but provide limited support for refinement and evaluation. Based on these findings, we identify nine design opportunities, including advanced visual interaction, simplified parameter control, precision editing, direct manipulation, workflow integration, default settings that support rapid exploration, and user guidance for later stages. We contribute a framework for analyzing GenAI interfaces and actionable directions for designing more usable, creativity-supportive GenAI image systems.

Preprint. Accepted for publication in CHI '26 (Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems). The final published version will appear in the ACM Digital Library.

CCS Concepts

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

Keywords

generative AI, image generation, interaction methods, interface design, creativity support tools

ACM Reference Format:

Hyerim Park, Malin Eiband, Andre Luckow, and Michael Sedlmair. 2026. Interaction Methods in Generative AI Image Tools: A Review of Trends and Design Opportunities Across HCI and Industry. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 27 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Generative AI (GenAI) image generation tools[1] enable users to produce high-quality visual content from diverse inputs (e.g., text, images, and sketches) without requiring advanced design or artistic expertise. Popular systems such as Midjourney [67], DALL·E [74], Firefly [2], and DreamStudio (built on Stable Diffusion [39]) [25] have rapidly gained traction in recent years. For example, the Midjourney subreddit [90] has over 1.8 million members, highlighting strong adoption among both hobbyists and professionals. This widespread adoption has drawn increasing interest from HCI researchers in addressing interaction challenges [13, 87, 131]. Although many GenAI tools support multimodal inputs, text prompts remain the primary interaction method and are often challenging to craft effectively [30, 57, 79]. As a result, prior HCI research has proposed prompt recommendation systems [17, 124], text input refinement tools [9, 103], and interfaces that combine visual and textual inputs [18, 82, 103]. However, a comprehensive synthesis of how GenAI image tools are designed to support user interaction remains scarce. We argue that building more usable and effective interfaces requires a systematic understanding of existing interaction strategies and their roles within creative workflows.

Motivated by this gap, we conducted a systematic review of GenAI image tool interfaces. We narrowed our scope to image generation due to its increasing importance in design practice [53, 79, 110] and the interaction challenges specific to this modality, such as aligning visual outputs with users' creative intent [18, 57, 78]. Prior work shows that professional designers often require more visually grounded input methods [55, 77, 79, 103], and that GenAI image tools demand interaction strategies distinct from those used in text generation, particularly for creativity support, visual refinement, and iterative workflows [116, 123]. By centering our review on GenAI image tools, we provide a focused and in-depth analysis of this rapidly evolving subset of GenAI applications [62, 98]. Our review spans both academic and commercial systems to capture emerging research directions alongside current practices. Academic tools were identified through a PRISMA process, while commercial tools were selected based on industry reports and prior HCI studies. In total, we analyzed 28 academic systems and nine commercial systems. To the best of our knowledge, this work presents the first systematic synthesis of GenAI image tool interfaces and interaction methods across both academic and commercial contexts, with a dedicated focus on image generation for design use. We address three research questions:

RQ1 What *interaction methods* are implemented in GenAI image tools in HCI research and industry?

RQ2 How do GenAI image tools support the *creative processes* in image generation?

RQ3 How can GenAI image tools be categorized by *functionality*, and how do these categories relate to interaction methods and the creative processes?

We conducted a thematic analysis of publications, tool documentation, and hands-on evaluations. Our analysis yielded three core frameworks: (1) *interaction methods*, categorized by input types (text, visual, attribute-based, and parameter control), control levels (global, component, and precision control), and interaction styles (direct manipulation, GUI interactions, parallel exploration, and iterative adjustments); (2) *creative processes*, including ideation, exploration, refinement, and evaluation; and (3) *tool functionalities*, including generators, editors, blenders, variation explorers, and prompt guides.

Our findings show that text prompts remain the dominant interaction method, while visual and attribute-based inputs are increasingly explored in HCI research. Commercial tools tend to emphasize parameter control, whereas academic systems investigate more semantic and visually grounded strategies, such as mood boards. Although most tools support early-stage creative activities such as ideation and exploration, they offer limited support for precise control, advanced refinement, and evaluation—capabilities that are critical for professional use. GUI implementations are typically basic, with few systems providing advanced features such as node-based or layer-based interfaces. In addition, prompt guidance remains limited in commercial tools, despite growing experimentation with guidance mechanisms in academic prototypes. Building on these findings, we identify nine design opportunities to improve visual and direct manipulation inputs, enhance parameter control mechanisms, integrate workflows across creative phases, and better support later-stage creative tasks. Our framework provides practical and actionable guidance for designing more usable, adaptable, and creativity-supportive GenAI systems.

2 Background

This section reviews prior work on GenAI image tools and their role in design workflows. We cover image generation models and tools, HCI research on GenAI interfaces, and situate these systems within the broader literature on AI-driven creativity support tools.

2.1 Generative AI for Image Generation

GenAI refers to AI technologies that produce synthetic content, including text, images, music, and videos. Models for image generation include Generative Adversarial Networks (GANs) [31], Variational Autoencoders (VAEs) [45], Denoising Diffusion Probabilistic Models (DDPMs) [36], Latent Diffusion Models (LDMs) [93], and Transformer-based text-to-image models [89]. Unlike earlier AI approaches in computer vision, which primarily focused on tasks such as classification, detection, or segmentation [47, 61, 91], GenAI emphasizes the generation of high-quality visual content that resembles human-created imagery [12, 26]. These capabilities have significant implications for creative fields such as design, where GenAI tools support ideation, productivity, and concept

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

development [6, 64, 71, 110]. Leading GenAI image tools such as Midjourney [67], DALL·E 3 [74], and Stable Diffusion platforms (e.g., DreamStudio, Automatic1111) [4, 25] support multimodal input and offer advanced features like inpainting and outpainting. Inpainting allows users to modify specific regions of an image by integrating new content into the existing context [73, 130], whereas outpainting extends images beyond their original boundaries while maintaining visual consistency [75, 129]. Despite these advances, most tools still rely primarily on text prompts, making effective prompt formulation critical for controlling image generation outcomes [55, 79, 99].

2.2 HCI Research on Generative AI Image Tool Interfaces

As GenAI tools gain wider adoption, HCI research has examined how these systems can be integrated into creative workflows while also addressing usability, ethical, and societal concerns [41, 53, 79, 110]. Prior work has identified key interaction challenges, including difficulties in interpreting generated outputs [16, 27, 30], frequent corrective interactions that reduce efficiency [37, 120], and steep learning curves associated with crafting effective prompts [79, 110]. Although GenAI tools can automate repetitive tasks, their limited contextual understanding often disrupts iterative workflows and leads to frequent tool switching [16, 79, 112]. Prompt creation remains a prominent usability barrier in GenAI image tools [58, 79]. Users frequently struggle to structure, refine, and evaluate prompts to achieve desired outputs [63, 79]. To address these challenges, researchers have proposed solutions such as prompt templates, keyword structuring strategies, and syntax-based formats (e.g., "[subject] in the style of [style]") [13, 81]. Some systems further implement prompt adaptation mechanisms that automatically refine user inputs to better align with users' intended goals [35].

Beyond text, visual inputs such as reference images paired with prompts have been shown to improve output relevance and accuracy [87]. Tools like PrompThis [34] support prompt–image comparison and history tracking for iterative refinement, while visual editors such as PromptChainer [128] provide node-based interfaces for constructing prompts, improving accessibility for novice users. In parallel, user communities play a central role in promoting effective usage of GenAI tools. Platforms such as r/Midjourney [90], PromptHero [86], and Lexica [52] facilitate the sharing of prompt strategies and results, while Discord communities for Midjourney [67] enable real-time feedback and collaborative learning.

Despite the growth of community resources and interface developments, comprehensive reviews focusing on GenAI **image** tool interfaces remain limited. Such synthesis is critical for identifying common design patterns, uncovering usability gaps, and informing future interface designs that better support creative workflows. Several prior surveys have examined GenAI tools across modalities [62, 76, 98, 123], covering domains such as text, image, and music generation, chatbots, and coding assistants. While these broader reviews include image-generation tools, they do not offer a focused synthesis of image-specific interaction methods or visual creative workflows. Our review addresses this gap by consolidating research on GenAI image tool interfaces and highlighting interaction patterns central to image-based creative practice.

2.3 AI-Driven Creativity Support Tools

In HCI, research on Creativity Support Tools (CSTs) has long examined systems designed to enhance human creativity across diverse domains. CSTs are defined as tools that "run on one or more digital systems, encompass creativity-focused features, and positively influence users of varying expertise across one or more phases of the creative process" [28, 101]. With the emergence of AI-driven CSTs, researchers have explored how intelligent systems can contribute more autonomously to creative tasks. Unlike traditional CSTs, which typically lack autonomy, AI-driven CSTs can perform tasks independently, introducing new roles and interaction paradigms. These systems support artistic vision, ideation, implementation, and evaluation, while also broadening accessibility to diverse user groups [19]. Hwang [40] observed that AI-driven CSTs may partially replace users during certain creative phases, making them particularly appealing for novice users. While such systems often excel at idea generation and execution, they provide limited support for very early-stage activities such as problem formulation or research. To address this imbalance, Hwang categorized AI-driven CSTs into four functional types: Generators (create content), Blenders (combine ideas), Editors (refine outputs), and Transformers (adapt content to new goals or styles). However, this framework was developed for general AI applications and does not fully capture the interaction characteristics of GenAI image tools. Verheijden and Funk [116] further noted that many GenAI image systems adopt a single-user, outcome-oriented model, limiting their applicability in collaborative or exploratory design workflows. Although their analysis offers valuable insights into GenAI image systems, it focuses on a limited set of tools and does not provide a comprehensive view of interaction patterns across the broader design landscape.

To date, systematic reviews of GenAI tools spanning both academic and commercial domains remain limited, particularly for image generation, where interaction methods and creative support are central concerns. This work addresses that gap by introducing analytical frameworks to identify design patterns, examine interaction challenges, and surface actionable opportunities for advancing GenAI image tool interfaces and interaction design.

3 Methodology

We adopted a two-step process to construct a comprehensive dataset for analyzing GenAI image tool design:

(1) **Systematic Review of HCI Tools:** We examined GenAI systems from peer-reviewed HCI studies, focusing on interaction methods, supported tasks, and functional roles.
(2) **Evaluation of Commercial Tools:** We analyzed widely used platforms to capture recent advances in real-world applications and their interface design.

This combination of academic and commercial sources reflects the evolving landscape of GenAI image generation tools across research and practice.

3.1 Academic Tools Selection

For the academic tool review, this study follows the PRISMA framework [68], a widely used approach in HCI for ensuring transparency

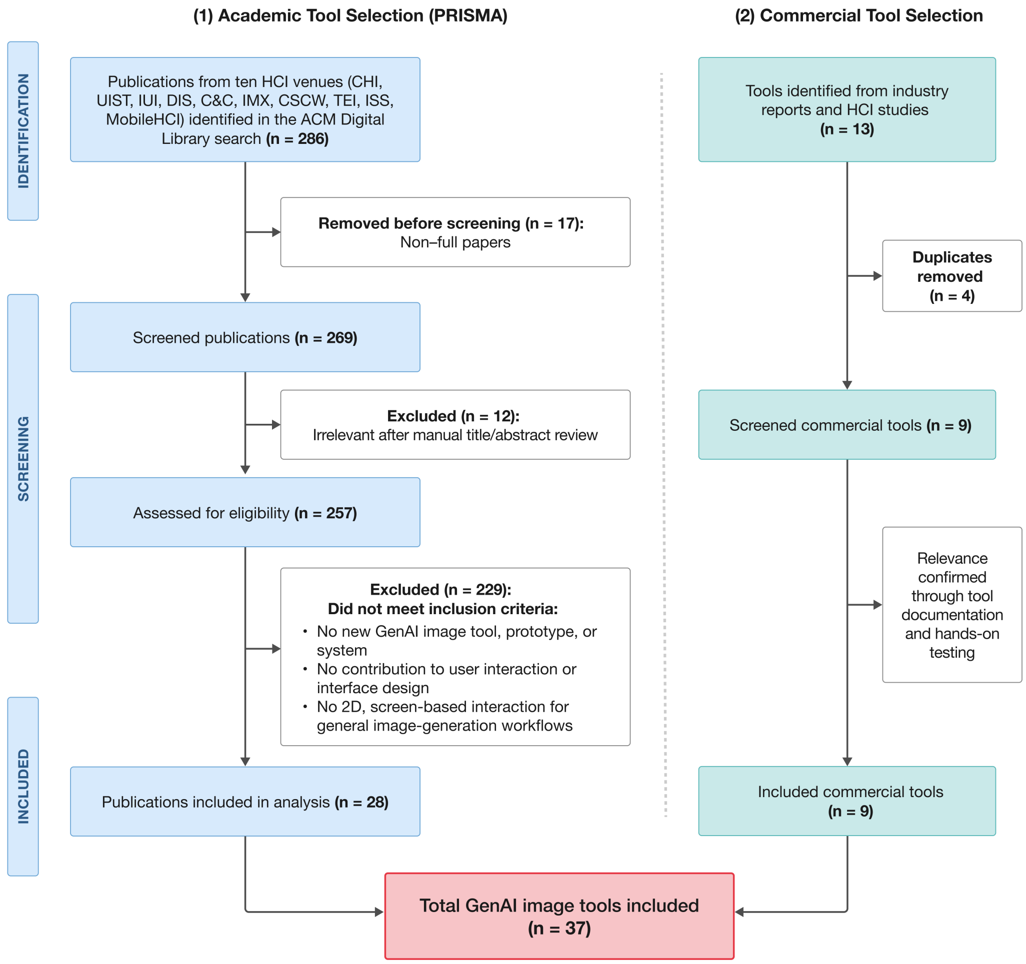


Figure 2: PRISMA flow diagram illustrating the selection process for academic and commercial GenAI tools.

and reproducibility. The subsections below outline the venue selection, search strategy, screening steps, and the resulting toolset (see Figure 2).

3.1.1 Venue Selection and Pilot Screening. Our review targeted papers introducing **new GenAI image tools** with a clear focus on **user-facing interaction and interface design**. Based on this scope, we first defined the publication venues to screen. The process began with the 28 SIGCHI-sponsored HCI conferences[2], which span a broad range of HCI research. After reviewing venue scopes, we excluded those clearly unrelated to our inclusion criteria (e.g., AutoUI, HRI, RecSys). Venues such as IMX, CSCW, TEI, ISS, and MobileHCI were retained, as they frequently publish interactive systems and creativity support tools, even when their primary focus lies in areas such as collaboration or tangible interfaces. For borderline venues—those partly relevant to interaction design but primarily oriented toward other topics (e.g., technical frameworks, conversational agents, or child-specific populations)—we conducted **pilot screenings** using our PRISMA search queries (January 2022–July 2025). We also pre-screened four technically oriented venues—TOG, ICML, CVPR, and ACM Multimedia—as representative samples from computer graphics, machine learning, computer vision, and multimedia applications. None yielded eligible papers, reflecting their topics outside of the focus on user interaction design. Full results and rationales are provided in the supplementary material. This process balanced breadth with focus, aiming to include relevant systems without unnecessarily expanding the review pipeline. The final set of ten venues included **CHI, UIST, IUI, DIS, C&C, IMX, CSCW, TEI, ISS, and MobileHCI.**

3.1.2 Identification Process.

Scope Definition. Our review focuses on **2D, screen-based GenAI image tools** published between **January 2022 and July 2025**. We set 2022 as the starting point because diffusion- and transformer-based text-to-image systems (e.g., DALL·E 2, Midjourney, Stable Diffusion) became widely accessible at that time [88, 93, 95] and rapidly entered creative practice [24, 46, 104, 113], prompting a

[2] https://sigchi.org/conferences/ (last accessed 25 January 2026)

surge of HCI work on their interaction challenges [53, 69, 79, 114]. We restrict our scope to 2D, screen-based interfaces, as they represent the dominant medium in current design workflows across domains. Moreover, they constitute the primary context in which GenAI-supported creative practices have been examined in recent empirical studies [17, 30, 41, 53, 79, 110]. This scope excludes 3D and spatial interfaces (e.g., AR/VR/MR) [5, 80, 135], which rely on fundamentally different interaction techniques such as spatial navigation, embodied manipulation, and viewpoint-dependent control [7, 8, 92]. We further include only **implemented, user-facing systems** with explicit interface contributions, excluding theoretical, qualitative-only, backend-only, or single-category tools (e.g., face-only generators) that do not generalize to open-ended image-generation workflows. To contextualize these scope decisions, we conducted two small contrast-set analyses—one for pre-2022 work and one for spatial modalities—which are summarized in Appendix A.1.

Search Strategy. All selected venues were indexed in the ACM Digital Library (ACM DL), and search queries were constructed following the ACM DL's advanced search guidelines[3]. Keywords such as "generative AI," "image generation," "interfaces," "HCI," "design," and "creative tool" were combined using Boolean operators and wildcards to ensure broad coverage while excluding purely theoretical or technical work. Although our review timeframe begins in 2022, we also included "GAN" as a search term to capture user-facing or hybrid tools that continued to employ GAN-based methods during this period [29, 36, 98, 116]. We focused on full research papers, which typically provide detailed descriptions of system design, implementation, and interaction goals. To ensure broad yet relevant retrieval, we used two complementary queries: a **title query** (broad, capturing papers that reference GenAI image tools even when interaction design is not explicitly mentioned), and an **abstract query** (narrower, filtering for explicit user interaction or interface design). Papers were included if they matched either query. Full query strings are provided in Appendix A.2. The initial search returned 286 papers; 17 short papers still appeared in the results despite our filters and were excluded under our eligibility criteria, leaving 269 for screening.

3.1.3 Screening Process. All 269 papers were screened using predefined inclusion and exclusion criteria by reviewing their titles and abstracts. When eligibility was unclear, we consulted the full text, with particular attention to introductions and system descriptions. Given our goal (RQ1–RQ3) of analyzing concrete interaction and interface patterns in implemented GenAI image tools, we prioritized systems that provided functioning prototypes and user-facing interaction designs.

Inclusion Criteria. Papers were included if they met all of the following criteria: (1) introduced a **new GenAI image tool, prototype, or system** designed and implemented by the authors; (2) contributed to **user interaction or interface design**, rather than focusing solely on model performance; and (3) employed **2D, screen-based interactions** applicable to general image-generation workflows.

[3]https://dl.acm.org/search/advanced (last accessed 25 January 2026)

Exclusion Criteria. Papers were excluded if they met any of the following criteria: (1) addressed **non-image GenAI domains** (e.g., text, music, audio); (2) presented **theoretical, conceptual, or guideline-only** work without an implemented system; (3) evaluated **existing tools** without introducing new interaction designs; (4) focused on **narrow domain contexts** with limited generalizability to creative workflows (e.g., education, clinical tools); (5) were **qualitative studies or literature reviews** without tool implementation; (6) focused **only on technical aspects** (e.g., model architectures, training procedures, or benchmarks) without user-facing interaction design contributions; (7) discussed **ethical, organizational, or social issues** without practical design implications; (8) focused on **3D generation or spatial modeling**; (9) focused on **face-generation–only** tools that do not generalize to open-ended image-generation workflows; (10) relied primarily on **spatial, tangible, or embodied interaction modalities** (e.g., AR/VR/MR, physical installations), which fall outside our focus on 2D, screen-based tools.

Borderline Decisions. Domain-oriented systems (e.g., fashion [127], interior design [121], character design [51]) were only included when their interaction methods generalized to broader GenAI image-generation workflows. For example, tools for physical product design [133] were included only if their primary interactions were 2D, screen-based, and interface-driven, making them applicable to general image-generation processes rather than dependent on tangible or spatial modalities.

Inter-rater Agreement. The screening process was conducted by the first two authors based on the predefined inclusion and exclusion criteria. The level of agreement between the two raters was assessed using Cohen's Kappa coefficient. Out of the 286 papers in the query, 278 (97.2%) were rated in agreement, while the number of agreements expected by chance was 233.9 (81.77%), yielding $\kappa = 0.847$ ($SE_\kappa = 0.053$, 95% CI [0.743, 0.951]). According to commonly used interpretations [49], this value indicates almost perfect agreement between the two raters beyond what would be expected by chance. All remaining discrepancies were discussed until a consensus was reached.

3.1.4 Final Tool Set. The screening identified 28 eligible tools, reflecting the emerging nature of GenAI image systems: one in 2022, five in 2023, 11 in 2024, and the remainder in 2025 (through July). Although this number is relatively modest, the tools exhibit diverse interaction methods and design features, providing a strong foundation for analysis. Our focus is on reviewing the **tools** themselves rather than the broader literature. The complete list is provided in Appendix A.4.

3.2 Commercial Tools Selection

Commercial tools were identified through two complementary sources. First, **industry data** on image generation volume (2022–August 2023) [43] and global market share in 2023 [106] were used to identify six leading platforms. Second, **HCI interview studies** with professional designers [15, 53, 78, 79, 110, 114, 131] surfaced GenAI image tools used in design practice. After removing four duplicates, we finalized nine distinct systems, verified through official product documentation and websites: DALL·E 3 (via ChatGPT) [74], Midjourney [67], Stable Diffusion [39], Adobe Firefly [2], Vizcom [118], RunwayML [94], VisualElectric [117], StarryAI [105], and NightCafe [70]. To ensure consistency across tools, we reviewed DreamStudio, the commercial interface built on Stable Diffusion, rather than open-source variants (e.g., Automatic1111 Web UI). For Midjourney, we analyzed the widely used Discord-based interface (launched in 2022) instead of the newer web version (July 2024). RunwayML was evaluated only for its image generation modules, excluding video, audio, and 3D features.

3.3 Analysis and Evaluation

3.3.1 Framework Development. We analyzed academic tools through their papers, examining introductions, design goals, system descriptions, and design guidelines, and also reviewed prototype videos when available. For commercial tools, we drew on documentation, tutorials, demos, and hands-on testing. The first author conducted the primary analysis, which was iteratively refined through discussions with the co-authors. For RQ1 and RQ2, we applied thematic analysis [10, 11]. Through open coding, we identified concepts such as interaction methods, input types, user control, and creative phases in image generation. These codes were organized and tagged using Notion tables [72] and then refined into overarching themes that formed the frameworks for *interaction methods* (RQ1) and *creative processes* (RQ2). For RQ3 (*tool functionalities*), we built on Hwang's framework [40] to capture functionalities specific to GenAI image tools by adding categories such as *Variation Explorer* and *Prompt Guides*. Unlike the original framework, which spans multiple AI domains (e.g., text and code), ours focuses exclusively on GenAI image generation.

3.3.2 Tool Evaluation and Scoring. Each tool was evaluated independently across the relevant framework dimensions—interaction methods, creative processes, and tool functionalities—using a four-level scale ranging from **strong (3)** to **none (0)**. In this context, "support" refers to whether a tool provides an implemented interaction capability within a given dimension. Because tools may support multiple dimensions simultaneously, categories in our framework are not mutually exclusive; mixed or cross-cutting support was therefore coded independently for each dimension. Scores reflect the presence, functional depth, and adaptability of interaction features (e.g., flexibility in user control or customization):

- **Strong (3):** Advanced functionality with adaptable options and relatively high usability.
- **Moderate (2):** Functional support with notable limitations (e.g., reduced adaptability or scope).
- **Limited (1):** Basic or indirect functionality with constrained interaction.
- **None (0):** No support for the interaction method.

Evaluations were iteratively refined through group discussions to ensure consistency, with representative examples documented to justify support levels (see supplementary material). The supplementary material provides category-specific criteria, complete scoring details showing how the 0–3 levels were operationalized, and supporting rationales, offering transparent and independently traceable evidence for all coding decisions and score assignments.

4 Results

We present our findings through three analytical frameworks—*interaction methods*, *creative processes*, and *tool functionalities*—each addressing one of our research questions (see Figure 1). The following sections summarize the key patterns observed, with full scoring details and rationales provided in the supplementary material. We illustrate how each framework is applied using *Brickify* [99] in Figure 4. Additional examples showing how we assigned scores (3–1) across all framework dimensions are provided in Appendix A.5.

4.1 RQ1: What *Interaction Methods* Are Implemented in GenAI Image Tools?

Interaction methods were analyzed across three dimensions: *input types*, *control levels*, and *interaction styles*. Overall tool support across these dimensions is summarized in Table 1.

4.1.1 Input Types. Input types refer to the different forms of information that users can provide to guide image generation. We identify four input types: *text*, *visual*, *attribute-based*, and *parameter control*. Figure 3 shows example tools representing each support level (3–1) across these four input types.

Text-Based Input. Text-based input, where users guide image generation with natural language, was the most common modality, strongly supported by 28 tools. Most systems provide free-form fields for detailed descriptions (e.g., subject, style, mood), while some restrict inputs to short keywords or predefined phrases [17, 55, 119]. Several tools assist prompt creation by refining vague inputs into concrete descriptions [9, 103] or by suggesting attributes such as style and tone from user-provided text or images [14, 17, 59, 111]. Others enhance usability through keyword suggestions as users type [117]. Interface layouts typically place text fields at the top, side, or bottom of the canvas, reserving the center for outputs or mood boards [14, 82, 122]. Commercial tools like DALL·E 3 and Midjourney adopt chat-based workflows, with Midjourney using command-driven prompts (e.g., `/imagine`) followed by a natural-language prompt. When tools support localized editing, text input is often paired with inpainting methods (e.g., masking, erasers, lassos). Firefly, WorldSmith [22], and Brickify [99] enable region-specific edits by linking text prompts to selected areas or objects.

Visual-Based Input. Visual input allows users to guide image generation by uploading reference images or sketches, or selecting predefined visual elements from libraries. This can reduce reliance on text prompts, which are often difficult to articulate precisely [48, 57, 79, 81]. Thirteen of the 37 tools provided strong support, offering diverse visual options and advanced menus that foreground visual thinking. Firefly, for instance, combines attribute controls (e.g., style, lighting, camera angles) with thumbnail previews and integrates generated images into traditional design workflows for adjustments such as color, lighting, or text style. Several systems emphasize sketches or drawings as primary input, ranging from basic pencil tools [51, 55, 125, 127] to more advanced brushes, lassos, and opacity controls [18, 56, 118]. SketchFlex [56] employs color-blocking to define spatial layouts, while Brickify [99] converts reference image elements into manipulable tokens. FusAIn [83] supports pen-based prompt composition by loading pens

Table 1: Support levels (strong, moderate, limited, none) across 37 GenAI image tools for the interaction methods framework. Tool counts appear in parentheses, followed by citations. Most tools strongly support text-based input and global control, whereas parameter control, precision control, and direct manipulation are less prevalent. † denotes borderline cases where user-facing semantics diverge from model-level implementation; see Appendix A.3 for the decision rubric.

	Interaction Methods	Strong Support	Moderate Support	Limited Support	No Support
Input Types	Text-Based	(28) [1–3, 9, 14, 18, 22, 25, 38, 51, 59, 66, 67, 70, 74, 82, 83, 94, 103, 105, 107, 111, 117, 118, 122, 124, 125, 133]	(5) [17, 55, 56, 99, 119]	(1) [121]	(3) [23, 126, 127]
	Visual-Based	(13) [2, 17, 18, 51, 55, 56, 83, 99, 103, 118, 125–127]	(12) [14, 22, 25, 67, 70, 82, 94, 105, 117, 119, 124, 133]	(7) [23, 38, 74, 107, 111, 121, 122]	(5) [1, 3, 9, 59, 66]
	Attribute-Based	(7) [2, 23, 38, 66, 67, 111], [18]†	(18) [1, 14, 17, 55, 56, 59, 70, 82, 83, 99, 105, 107, 117–119, 122, 124], [121]†	(9) [3, 9, 25, 51, 74, 94, 103, 125, 133]	(3) [22, 126, 127]
	Parameter Control	(1) [67]	(5) [25, 51, 70, 94, 105]	(17) [2, 3, 9, 18, 23, 55, 56, 66, 74, 83, 117–119, 121, 124, 125, 133]	(14) [1, 14, 17, 22, 38, 59, 82, 99, 103, 107, 111, 122, 126, 127]
Control Levels	Global Control	(29) [1–3, 9, 14, 17, 18, 22, 25, 38, 51, 55, 56, 59, 66, 67, 70, 74, 82, 83, 94, 99, 103, 105, 111, 117, 122, 124, 125]	(4) [107, 118, 121, 133]	(4) [23, 119, 126, 127]	(0) –
	Component Control	(12) [2, 18, 22, 51, 56, 82, 83, 99, 118, 124, 125, 133]	(15) [14, 25, 38, 55, 67, 70, 74, 94, 103, 105, 107, 117, 122, 126, 127]	(6) [3, 17, 23, 111, 119, 121]	(4) [1, 9, 59, 66]
	Precision Control	(3) [2, 83, 118]	(3) [18, 51, 133]	(17) [22, 25, 55, 56, 67, 70, 74, 82, 94, 99, 103, 105, 117, 124–127]	(14) [1, 3, 9, 14, 17, 23, 38, 59, 66, 107, 111, 119, 121, 122]
Interaction Styles	Direct Manipulation	(1) [99]	(10) [2, 18, 22, 51, 55, 56, 83, 118, 125, 133]	(16) [9, 17, 23, 25, 67, 70, 74, 82, 94, 103, 105, 117, 119, 124, 126, 127]	(10) [1, 3, 14, 38, 59, 66, 107, 111, 121, 122]
	GUI Interactions	(10) [1, 2, 9, 14, 18, 38, 66, 107, 118, 124]	(22) [3, 17, 22, 23, 25, 51, 55, 56, 59, 70, 82, 83, 94, 99, 103, 105, 111, 117, 121, 122, 125, 133]	(5) [67, 74, 119, 126, 127]	(0) –
	Parallel Exploration	(19) [3, 9, 14, 17, 22, 23, 51, 59, 66, 82, 103, 107, 111, 117–119, 121, 122, 133]	(14) [1, 2, 18, 25, 38, 55, 56, 67, 70, 83, 94, 99, 105, 124]	(4) [74, 125–127]	(0) –
	Iterative Adjustments	(9) [2, 18, 22, 51, 83, 99, 118, 124, 133]	(20) [3, 14, 25, 55, 56, 66, 67, 70, 74, 82, 94, 103, 105, 107, 111, 117, 122, 125–127]	(8) [1, 9, 17, 23, 38, 59, 119, 121]	(0) –

with attributes like texture or color for fine-grained control. Visual search is another common feature. GenQuery [103] enables refinement through gallery-based selection. Beyond search, tools also support visual exploration and manipulation: PromptPaint [18], for example, introduces a palette metaphor for mixing colors or adding stencils and layers. Tools further support visual brainstorming through node-based interfaces [14], scene graphs [38], mood boards [17, 82, 119], and layout panels [38]. Some systems offer only limited visual input, such as simple reference-image uploads or theme/style folders [22, 67, 133]. Commercial tools commonly extend visual input through style libraries, palettes, and builder interfaces [25, 70, 94, 105, 117]. Five tools lacked visual input entirely, relying exclusively on text-based prompts [1, 3, 9, 59, 66].

Attribute-Based Input. We classify controls as attribute-based when they adjust higher-level semantic (e.g., style, theme) or visual properties (e.g., lighting, color, composition, layout), rather than exposing explicit model- or inference-level settings. This classification is based on **user-facing semantics**: controls presented as specifying aesthetic, semantic, or spatial properties are treated as attribute-based even when implemented through model-level mechanisms such as ControlNet [60]. For instance, RoomDreaming [121] uses ControlNet depth and segmentation internally to guide *layout coherence*, but exposes these adjustments through design-oriented controls (e.g., "new design directions," "like" feedback) rather than system parameters. PromptPaint [18] similarly presents prompt-weight adjustments via a palette-like interface; although these operate on attention weights internally, they are framed as semantic or stylistic controls. In contrast, NightCafe labels comparable sliders using system-level terms (e.g., seed, prompt weight), and we therefore classify them as parameter controls. Because attribute-oriented controls may rely on model-level mechanisms, classification ambiguity can arise; we resolve such cases using the decision rubric in Appendix A.3, which formalizes our operational criteria for borderline examples (e.g., ControlNet depth, prompt-weight sliders).

Seven tools offered strong attribute-based support, including attribute blending, weighting, and GUI-driven customization (e.g., sliders, dropdowns, toggles). These controls allow users to adjust semantic or visual properties such as textures, styles, materials, lighting, camera angles, or attribute-rich tokens extracted from reference images. Systems such as PromptPaint [18] and PlantoGraphy [38] support fine-grained semantic adjustments (e.g., "matte" vs. "glossy," time of day, weather), while Midjourney and SketchFlex [56] provide command-based (e.g., `–style raw`) or sketch-informed controls that steer stylistic and compositional attributes. Moderate support (18/37 tools) was observed in systems offering only fixed presets or limited stylistic options. Several domain-specific tools provide more specialized attribute sets—for example, sleeve length in fashion design, plant size in landscaping, or room arrangement in interior design—reflecting attributes tailored to particular applications [23, 38, 121].

Parameter Control. Parameter control exposes **settings at the model or inference level**, such as seed values, batch size, sampling steps, denoising strength, or prompt weight. These controls are typically framed in numeric or system-level terms, often requiring familiarity with underlying model behavior and thus potentially limiting accessibility [54]. Among the reviewed tools, only Midjourney offered comprehensive parameter controls (e.g., seed,

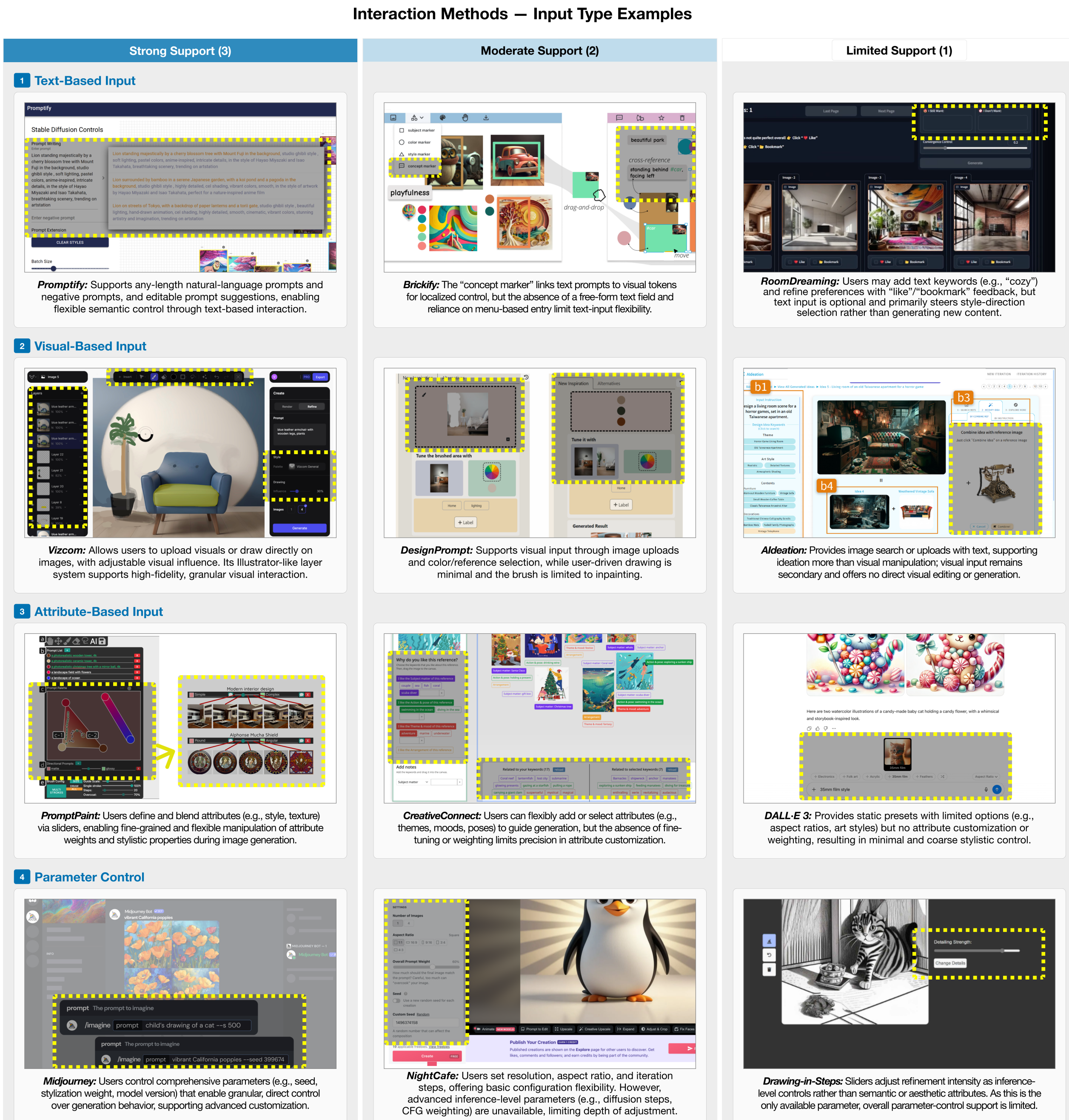


Figure 3: Example tools illustrating contrasting support levels (scores 3–1: strong, moderate, limited) for the *Input Types* dimension. A score of 0 (none) indicates feature absence and is therefore omitted from the figure. Each row presents representative tools for one input type—*text-based*, *visual-based*, *attribute-based*, and *parameter control*—showing how the scoring criteria distinguish between levels of support. Representative tools include Promptify [9], Brickify [99], RoomDreaming [121], Vizcom [118], DesignPrompt [82], AIdeation [122], PromptPaint [18], CreativeConnect [17], DALL·E 3 [74], Midjourney [67], NightCafe [70], and Drawing-in-Steps [125]. Additional examples (triads) for all framework dimensions appear in Appendix A.5. Images are adapted from ACM-published works (© ACM), CC BY 4.0 licensed sources, and screenshots of commercial GenAI interfaces (© Vizcom, © OpenAI for DALL·E 3, © Midjourney, © NightCafe), used for scholarly analysis and illustration.

image weight, model version, randomness). To avoid ambiguity with attribute-based controls, we classify settings as parameter control only when they are presented explicitly in numeric or system-level terms. For example, Midjourney's `–seed 123`, `–chaos 30`, and model-version switches expose inference-level parameters rather than semantically framed controls (see Appendix A.3). A few commercial tools, including DreamStudio and RunwayML, supported moderate parameter tuning (e.g., sampling steps, image count) through expandable menus [25, 94, 105]. Most academic tools offered limited or no parameter control. Limited support (17/37 tools) typically exposed only one or two system-level settings, such as batch size, image count, attention weights, or prompt influence [9, 18, 117, 118, 124, 133], while 14 tools did not expose any model parameters at all.

4.1.2 Control Levels.

This dimension captures the granularity of user control in image generation, categorized into *global*, *component*, and *precision* control.

Global Control. Global control allows users to define overarching aspects of an image, such as theme, style, or composition, typically through high-level text prompts, dropdowns, or preset libraries. For example, a prompt like "a futuristic cityscape" produces cohesive results that implicitly determine layout, mood, and detail, which users can then refine using more specific descriptions. Most tools (29/37) offered strong global control, enabling users to generate coherent images from conceptual inputs such as style, tone, or theme. Several domain-specific systems also provided robust support: AutoSpark [14] supports car design variations through keywords (e.g., "futuristic," "fluid"), PlantoGraphy [38] provides landscape presets, and Paratrouper [51] ensures stylistic consistency across character sets. Some tools offered only moderate support. For instance, Vizcom, which focuses on 3D product renderings, accepts conceptual keywords but relies on detailed prompts or sketches for coherence. Similarly, ProtoDreamer [133] and RoomDreaming [121] require initial image uploads, limiting flexibility. Tools lacking global control typically depend on random generation or rigid templates [119, 127].

Component Control. Component control enables users to modify specific elements, such as shapes, materials, or object arrangements, while preserving the overall composition. For example, prompts like "Change the green couch to a red vintage couch" or "Remove trees in the background" target individual components without requiring fine-grained editing. This level of control often involves masking regions and issuing localized prompts. Some domain-specific tools, such as AutoSpark [14], lack inpainting but allow targeted regeneration of car parts. PlantoGraphy [38] uses graph-based inputs for plant placement, offering partial control over spatial relationships. Other tools provided only limited support: edits were possible but often affected unrelated elements or lacked precision [23, 119]. Four tools offered no component-level editing at all, relying instead on global generation or broad adjustments via prompts, sliders, or navigation elements [1, 9, 59, 66].

Precision Control. Precision control supports fine-grained adjustments to specific attributes, such as regenerating localized regions, resizing objects to exact dimensions, or adjusting spacing by defined percentages. Supporting this level of control requires tools that can accurately interpret and execute detailed user modifications. We evaluated precision control based on a tool's ability to support such granular changes, including refining textures, editing small regions, and customizing detailed elements. Although inpainting with text prompts provides some degree of precision, these methods can misalign with user intent, making edits time-consuming and error-prone. Firefly and Vizcom demonstrated strong precision support, enabling refined edits to materials, lighting, and textures, as well as sketch- or palette-based revisions. FusAIn [83] also supports localized pen-based edits of attributes and regions. Among the evaluated tools, 14 lacked precision control, 17 offered only limited support via inpainting or text-based editing, and one tool, ProtoDreamer [133], enabled anchor-point adjustments for more accurate object resizing.

4.1.3 Interaction Styles.

Interaction styles describe how users engage with inputs or interface elements in GenAI image tools. We identify four approaches: *direct manipulation*, *GUI interactions*, *parallel exploration*, and *iterative adjustments*.

Direct Manipulation. Direct manipulation lets users interact with visual content by dragging, resizing, or drawing on images, mimicking interactions with physical materials. This approach enhances engagement and accuracy through immediate feedback, clear visuals, and reversible actions [100, 102]. Across the reviewed tools, three forms of direct manipulation were observed. First, inpainting tools such as brushes, erasers, or selection tools paired with text prompts enabled localized edits [25, 67, 74, 82, 105, 124]. Second, some tools supported element-level interaction, allowing users to drag, drop, or resize components. For instance, Brickify [99] manipulates visual tokens, and FusAIn [83] supports pen-based edits of attributes or regions. Third, systems such as mood boards and grid layouts allowed images to be grouped, resized, or repositioned as whole units rather than edited individually [9, 17, 23, 82, 103, 119]. These whole-image interactions were therefore rated as limited in direct manipulation. Moderate support was observed in tools offering advanced selection mechanisms, such as layer-based editing [18, 118], draggable palette components [2], or region-based resizing with anchors [133]. Some systems also enabled localized refinement through color-coded segmentation maps [22]. Only one tool, Brickify [99], demonstrated strong object-level manipulation; most tools lacked direct editing of individual components in generated images.

GUI Interactions. GUI interactions let users adjust attributes through graphical elements such as sliders, buttons, and dropdowns. These controls refine outputs without directly manipulating content. While sometimes grouped under direct manipulation [65, 102], we treat them separately to emphasize their role in tuning attributes or parameters rather than directly editing visual content. Ten tools offered strong support for GUI interactions, providing advanced controls beyond basic sliders or buttons. Examples include prompt weight mixing [18], semantic adjustments such as weather, time, or art style [38, 118], and visual effects like lighting or camera angles [2]. Some systems updated labels or options dynamically based on user input or generation history [9, 124], while others added layout navigation or history tracking via scene graphs or node maps [14, 17, 82]. DreamSheet [3] further expanded this category

Table 2: Support levels (strong, moderate, limited, none) for 37 GenAI image tools across four creative process phases. Tool counts appear in parentheses, followed by citations. Most tools emphasize early-stage activities such as ideation and exploration, while refinement and evaluation receive less support.

Phase	Creative Focus	Strong Support	Moderate Support	Limited Support	No Support
Ideation	Generating new concepts (*breadth, novelty, inspiration*)	(31) [1–3, 9, 14, 17, 18, 25, 38, 51, 55, 56, 59, 66, 67, 70, 74, 82, 83, 94, 99, 103, 105, 107, 111, 117–119, 121, 122, 124]	(6) [22, 23, 125–127, 133]	(0) –	(0) –
Exploration	Experimenting with selected designs (*depth, variation, recombination*)	(28) [2, 3, 14, 18, 23, 25, 38, 51, 66, 67, 70, 74, 82, 94, 103, 105, 107, 111, 117–119, 121, 122, 124–127, 133]	(9) [1, 9, 17, 22, 55, 56, 59, 83, 99]	(0) –	(0) –
Refinement	Detailed editing and control (*precision, control, polish*)	(7) [2, 18, 56, 83, 99, 118, 133]	(19) [14, 22, 25, 51, 55, 66, 67, 70, 74, 82, 94, 103, 105, 111, 117, 124–127]	(11) [1, 3, 9, 17, 23, 38, 59, 107, 119, 121, 122]	(0) –
Evaluation	Comparing and selecting outputs (*judgment, comparison, selection*)	(3) [14, 66, 133]	(12) [1–3, 9, 22, 23, 51, 103, 107, 118, 121, 124]	(22) [17, 18, 25, 38, 55, 56, 59, 67, 70, 74, 82, 83, 94, 99, 105, 111, 117, 119, 122, 125–127]	(0) –

with a spreadsheet-style GUI in which prompts and parameters are edited in grid cells, offering a structured interface for generative control. Tools with moderate support typically included multiple GUI elements but offered limited flexibility. Many relied on static options or fixed parameters [22, 23, 59, 103, 121, 133], such as prompt weighting, image count selection, or preset menus, resulting in less control than tools with strong GUI support.

Parallel Exploration. Parallel exploration enables users to experiment with multiple inputs or design configurations simultaneously, fostering creativity through breadth and comparison. This typically involves generating image variations and offering interfaces for efficient exploration and side-by-side evaluation. Nineteen tools offered strong support, and 14 offered moderate support. Tools with strong support generated multiple variations from text prompts, keywords, or layouts, and enabled interactive comparison. Examples include mood board interfaces for rearranging and comparing outputs [9, 17, 119], design-thinking maps linking images with keywords [14], workbench-style layouts [118], and matrix views showing attribute variations [23, 133]. DreamSheet [3] extended this concept with a spreadsheet-style GUI in which each row represents a prompt–parameter combination, enabling structured and scalable generation. Tools with moderate support displayed variations in more constrained ways, such as history panels [38, 124] or small side views [2], rather than central, interactive spaces. Systems requiring manual re-entry of prompts for each variation, even when paired with otherwise strong interfaces, were also rated as providing moderate support [22]. Tools with limited support, such as DALL·E 3 (via ChatGPT), required scrolling through chat histories to manage outputs, reducing exploration efficiency.

Iterative Adjustments. Iterative adjustments refer to incrementally refining a selected image to achieve greater precision or specific creative goals. Tools that did not allow output selection for targeted refinement were rated as offering no support for iterative adjustments. Some tools supported recombination, keyword replacement [9, 17, 119], or image arrangement [17], but enabled only broad conceptual changes without preserving specific visual elements or supporting localized refinement. Two domain-specific tools for fashion design and landscaping provided limited support [23, 38]. While lacking regional selection, they allowed partial refinement through layout adjustments or sliders controlling properties such as sleeve length or color. Moderate support (20/37 tools) was observed in systems that enabled region-based refinement via inpainting and text prompts. Nine tools offered strong support by adding parameter-weighting sliders for selected regions [18, 133], layer systems, and diverse editing tools such as sketching or adding stickers or icons [2, 18, 118]. Some systems also included overview panels or history nodes for tracing and refining earlier outputs [22, 124].

4.2 RQ2: How Do GenAI Image Tools Support the *Creative Processes* in Image Generation?

Our analysis examined how GenAI tools support four phases of the creative process in image generation: *ideation*, *exploration*, *refinement*, and *evaluation*. Broader stages from creativity frameworks [40], such as problem finding, knowledge gathering, and finalization, were excluded because few tools in our dataset addressed them. Because image generation is rarely linear, users often move fluidly between phases [79]. Table 2 summarizes support levels across tools.

4.2.1 Ideation. Ideation represents a divergent phase of the creative process, where the goal is to generate new directions, promote **breadth**, and spark **inspiration**. Tools supporting this phase emphasize **variety**, **novelty**, and **idea generation** over precision. Most tools (31/37) strongly supported ideation by generating initial concepts and encouraging exploration of multiple directions. Many produced variations based on prompts or random generation (e.g., a "random" button) [119], supporting brainstorming and expansion of the design space [17, 111]. Some tools provided preset menus for styles, themes, or materials to reduce input effort and accelerate concept generation [2, 25, 70, 94, 105, 117]. Several systems enabled the combination of text and image inputs, reuse of generated outputs as new inputs, keyword replacement, or multimodal prompting to produce variation [17, 33, 82, 84, 103, 119, 124]. Others assisted ideation by recommending prompts in real time [9, 17] or by supporting prompt steering through commands or sliders (e.g., "creative" vs. "similar") [2, 67, 121]. Commercial platforms often leveraged large user bases and curated content to surface example prompts and images [67, 70, 105, 117], whereas HCI research prototypes favored structured interfaces such as mood boards, grids, or canvases for organizing and experimenting with outputs [9, 14, 17, 82, 111, 119, 121, 122].

4.2.2 Exploration. Exploration builds on ideation by shifting from breadth to **depth**, emphasizing **variation**, **recombination**, and **testing** of design directions, starting from selected images. Most tools (28/37) strongly supported this phase by enabling users to regenerate images with adjusted prompts [119, 122], mix attributes such as colors or effects [18], or rearrange layouts [38, 99, 121]. In contrast, tools limited to prompt modification without reference image selection or attribute control were rated as offering moderate support [9, 22, 59]. Tools that lacked the ability to branch from specific images were likewise rated as providing moderate support [1, 17]. Commercial systems frequently supported exploration through parameter tuning (e.g., prompt weights, CFG scales, or style modifiers such as `--stylized`), whereas HCI tools typically exposed fewer adjustable parameters but emphasized usability through GUI elements such as sliders, node graphs, or scene graphs [18, 23, 38, 121].

4.2.3 Refinement. Refinement focuses on polishing outputs with **precision**, offering users more **targeted control** than earlier phases. While many tools supported basic adjustments, only seven provided strong refinement capabilities [2, 18, 83, 99, 118, 125, 133]. These tools supported component-level editing (e.g., size, position, addition or removal of objects), combining direct manipulation techniques such as sketching, dragging, or inpainting with GUI controls, including sliders and toggles [2, 18, 99, 118, 133]. Advanced features included layer-based editing, granular inpainting, resizing anchors, and localized adjustment panels. For example, Firefly's "Generative Fill" enabled precise edits via GUI inputs integrated with Photoshop, while FusAIn [83] supported region-specific pen-based editing with custom attributes. Moderately supported tools [25, 67, 70, 74, 105, 117] offered inpainting but relied mainly on text prompts, with limited parameter control. Some tools also supported parametric adjustments, allowing users to fine-tune attributes with GUI sliders [2, 18] or numerical prompt inputs [67]. Domain-specific systems contributed specialized refinement, such as PlantoGraphy's scene graphs for plant placement [38]. Tools designed primarily for ideation or exploration (e.g., keyword recombination or minor variations) did not align with the goals of this phase [9, 17, 59, 122].

4.2.4 Evaluation. Evaluation emphasizes **assessing** and **comparing** outputs to determine alignment with creative goals. Unlike refinement, which focuses on modification, evaluation centers on **judgment**, **reflection**, and **selection**. Only three tools provided strong evaluation support. AutoSpark [14] enabled grouping by attributes (e.g., texture, shape, color) and linked keywords to regions via heatmaps. ProtoDreamer [133] used a matrix view to visualize the effects of prompt weights. Varif.ai [66] supported diversity evaluation through histograms, tooltips, and history panels, facilitating comparison and tracking of variation. Twelve tools, mostly academic and one commercial, provided moderate support. Examples include mood board interfaces for grouping and comparison [9, 17, 59, 103], and global views such as trees or history panels linking outputs to prompts [22, 124]. Most tools lacked structured comparison mechanisms, and evaluation often relied on manual inspection, with grid views limited to a small number of images (e.g., four) and designed more for rapid variation than deliberate selection.

Note on Finalization. Finalization, which involves image quality optimization (e.g., upscaling or resolution tuning), was excluded from analysis. While some commercial tools include such features [2, 67], they reflect general image editing rather than GenAI-specific creativity support. None of the reviewed HCI tools explicitly supported this phase.

4.3 RQ3: How Can GenAI Image Tools Be Categorized by *Functionality,* and How Do These Categories Relate to Interaction Methods and the Creative Processes?

Tools were categorized by core functionality, adapting Hwang's framework of *Editors*, *Transformers*, *Blenders*, and *Generators* [40]. While originally developed for AI tools more broadly, we tailored the framework to GenAI **image** tools. The *Transformers* category was excluded, as content conversion (e.g., text to code snippets) was not represented in our dataset. To address gaps in the original classification, we introduced two new categories: *Variation Explorers*, which help users navigate design possibilities beyond generation, and *Prompt Guides*, which support crafting and refining text prompts. Table 3 summarizes support levels across these categories.

Generators. Generators create new images from user input, forming the core capability that distinguishes GenAI systems from traditional AI tools (e.g., predictive analytics, classification, or automation) [96, 108, 109]. Most tools strongly supported this function. Two offered only moderate support, as they did not allow descriptive text [23] or reference image uploads [127], relying instead on predefined references.

Editors. Editors modify or enhance existing content through targeted changes, such as adjusting colors, layouts, or removing elements, often by regenerating localized regions while preserving the rest of the image. They act as refiners, focusing on localized modifications rather than global regeneration. Strong support was characterized by precise region selection, fine-grained controls, and accurate refinements.

Blenders. Blenders merge multiple images into a cohesive output. Some produce system-driven results, while others allow users to adjust blending parameters such as image weighting or output ratios. Although blending prompts is common, direct multi-image blending is less widely supported. Tools limited to style pickers (e.g., tone or theme menus) were not classified as blenders, as they modify overall aesthetics rather than combine inputs.

Variation Explorers. Variation Explorers guide users in systematically exploring design alternatives. Instead of producing single outputs, they organize variations through features such as sliders for style or mood, branching trees for design directions, and grids or galleries for attribute combinations. Unlike *Exploration* as a creative phase or *Parallel Exploration* as an interaction style, this category emphasizes tool functionalities that actively support guided experimentation and comparison across outputs.

Prompt Guides. Prompt Guides support prompt construction through suggestions, templates, and interactive guidance. Features

Table 3: Support levels (strong, moderate, limited, none) for 37 GenAI image tools across five functionality categories. Tool counts appear in parentheses, followed by citations. Generators and variation explorers are strongly supported, whereas editors, blenders, and prompt guides are comparatively less supported overall.

Categories	Strong Support	Moderate Support	Limited Support	No Support
Generators	(35) [1–3, 9, 14, 17, 18, 22, 25, 38, 51, 55, 56, 59, 66, 67, 70, 74, 82, 83, 94, 99, 103, 105, 107, 111, 117–119, 121, 122, 124–126, 133]	(2) [23, 127]	(0) –	(0) –
Editors	(7) [2, 18, 51, 83, 99, 118, 133]	(16) [22, 25, 55, 56, 67, 70, 74, 82, 94, 103, 105, 117, 124–127]	(8) [3, 14, 23, 38, 107, 111, 119, 122]	(6) [1, 9, 17, 59, 66, 121]
Blenders	(5) [2, 51, 67, 107, 118]	(7) [18, 23, 66, 83, 99, 119, 126]	(11) [3, 22, 82, 94, 103, 111, 117, 122, 125, 127, 133]	(14) [1, 9, 14, 17, 25, 38, 55, 56, 59, 70, 74, 105, 121, 124]
Variation Explorers	(28) [1–3, 9, 14, 18, 22, 23, 25, 38, 51, 59, 66, 67, 70, 82, 94, 103, 105, 107, 111, 117–119, 121, 122, 124, 133]	(9) [17, 55, 56, 74, 83, 99, 125–127]	(0) –	(0) –
Prompt Guides	(12) [1, 9, 14, 17, 56, 59, 82, 103, 107, 111, 122, 124]	(5) [55, 66, 99, 117, 119]	(12) [2, 3, 18, 23, 25, 38, 51, 67, 83, 94, 118, 125]	(8) [22, 70, 74, 105, 121, 126, 127, 133]

include keyword recommendations from user inputs, automatic refinements, and controls such as dropdowns or sliders. Unlike *Variation Explorers*, which focus on experimenting with outputs, Prompt Guides emphasize input construction, making prompt design more systematic and accessible.

4.4 Cross-Framework Relationships

To examine relationships between functionality, interaction methods, and creative processes, we computed Pearson correlations across 37 tools.[4] All heatmaps with statistically significant results ($p < .05$) under False Discovery Rate (FDR) correction ($q < .05$) are provided in Figure 5. Effect sizes follow Cohen's guidelines (small ≈ .10, medium ≈ .30, large ≥ .50), and we report r, p, and FDR-adjusted q values.

Consistent Alignments. Our analysis revealed several consistent correlations. *Refinement* correlated strongly with *Editors* ($r = .84$, $p < .001$, $q < .001$), and *Visual-Based Input* with *Editors* ($r = .75$, $p < .001$, $q < .001$). Early-stage patterns linked *Text-Based Input* with *Generators* ($r = .69$, $p < .001$, $q < .001$), *Generators* with *Ideation* ($r = .54$, $p < .001$, $q < .01$), and *Global Control* with *Generators* ($r = .61$, $p < .001$, $q < .001$). These findings align with prior work showing that GenAI tools primarily use text prompts for ideation, while visual input more often supports refinement [50, 78].

Emerging Patterns. Prompt Guides correlated positively with *Ideation* ($r = .47$, $p < .01$, $q < .05$) but negatively with fine-grained controls (*Parameter Control* $r = -.43$, $p < .01$, $q < .05$; *Precision Control* $r = -.41$, $p < .05$, $q < .05$). *Variation Explorers* showed positive correlations with the *Evaluation* phase ($r = .43$, $p < .01$, $q < .05$) and with *Parallel Exploration* ($r = .62$, $p < .001$, $q < .001$), indicating their close relationship to comparison and selection tasks, often supported through side-by-side grids or mood boards. Their association with the broader *Exploration* phase was weaker and only marginally significant ($r = .41$, $p < .05$, $q \approx .06$). *Variation Explorers* were negatively correlated with hands-on interaction methods such as *Direct Manipulation* ($r = -.43$, $p < .01$, $q < .05$) and *Visual-Based Input* ($r = -.49$, $p < .01$, $q < .01$). This contrasts with the strong positive alignment observed between *Visual-Based Input* and *Editors* ($r = .75$, $p < .001$, $q < .001$). Editor-focused tools were less likely to support *Parallel Exploration* ($r = -.36$, $p < .05$, $q < .05$). Evaluation benefited from structured interfaces: *GUI Interactions* ($r = .51$, $p < .01$, $q < .01$) and *Parallel Exploration* ($r = .48$, $p < .01$, $q < .01$) both aligned with *Evaluation*. Finally, *Blenders* correlated with *Precision Control* ($r = .45$, $p < .01$, $q < .01$) and *Iterative Adjustments* ($r = .45$, $p < .01$, $q < .01$).

Summary. Overall, *Text-Based Input* and *Global Control* aligned with *Generators* and supported ideation, reflecting the frequent use of prompts to initiate new ideas. *Editors*, often paired with *Visual-Based Input*, were closely associated with refinement, during which users focused on modification and localized changes. *Variation Explorers*, often combined with *Parallel Exploration*, supported evaluation through comparison and selection interfaces such as grids or mood boards, but were rarely integrated with direct manipulation. *Prompt Guides* also supported ideation by providing structured input assistance, yet showed negative correlations with *Precision Control*, suggesting that these features are more common in early stages and less often combined with fine-grained adjustments. Finally, *Blenders* correlated with *Precision Control* and *Iterative Adjustments*, indicating that compositing and merging are typically carried out through stepwise refinement later in the workflow.

To provide a high-level view of how tools are distributed across the analytical frameworks, we include a global Sankey diagram (Figure 6). For each tool, support levels were paired across adjacent frameworks using our 0–3 ordinal scale and converted into a co-support score (strong–strong = 3, strong–moderate = 2, moderate–moderate = 1, otherwise = 0). These scores determine flow widths, emphasizing meaningful co-occurrences while suppressing incidental links. The Sankey integrates the three frameworks into a single visualization, offering an at-a-glance summary of dominant and underrepresented pathways across the design space.

5 Design Space and Opportunities for GenAI Image Tool Interfaces

Our findings across *interaction methods*, *creative processes*, and *tool functionalities* reveal several system-level design imbalances in current GenAI image tools. As summarized in Table 1–3 and the correlation heatmaps in Figure 5, existing systems strongly emphasize

[4] Correlations describe co-occurrence patterns, not causal effects.

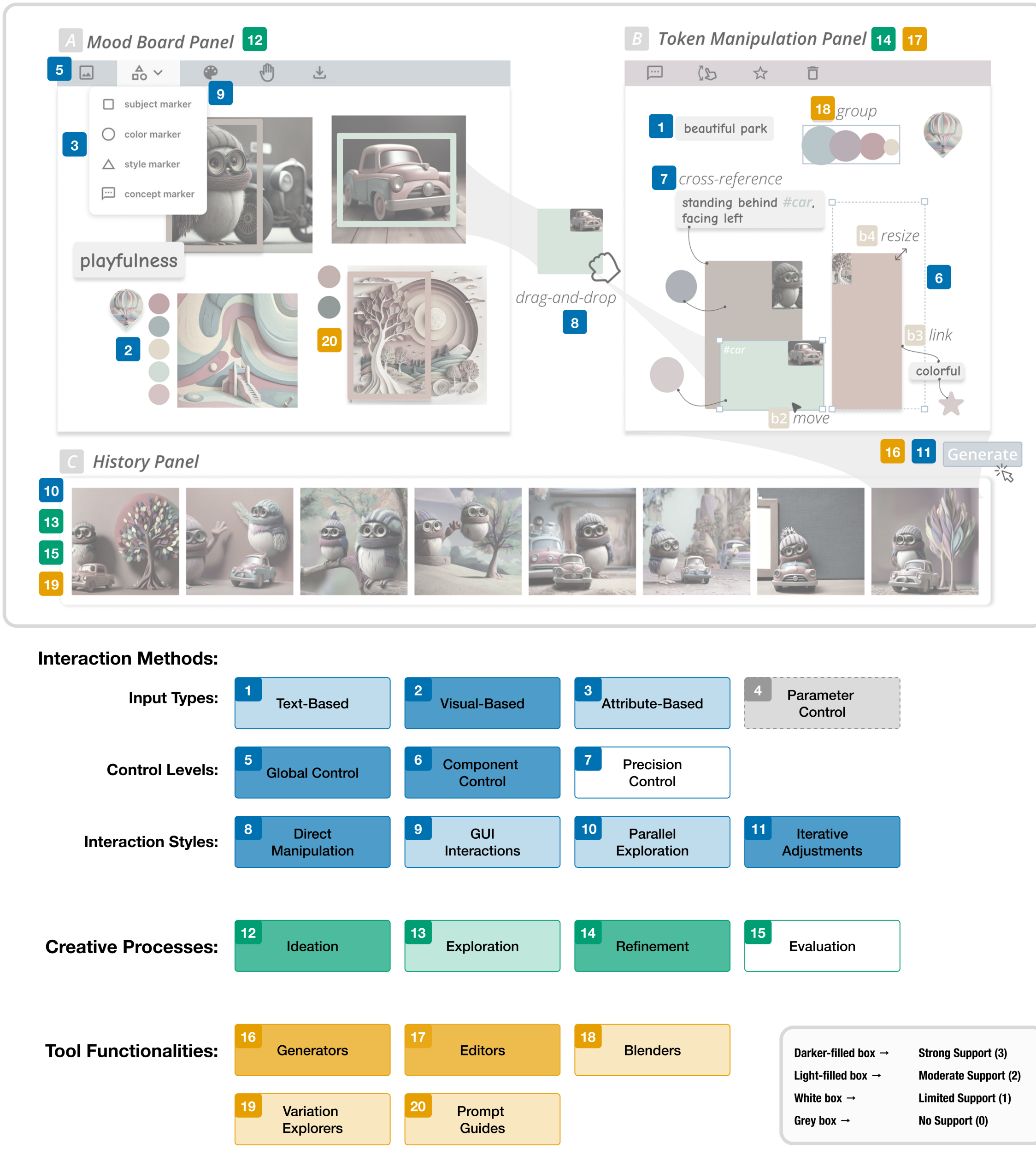


Figure 4: Application of our evaluation framework to *Brickify*. Figure adapted from Shi et al. [99], © ACM, used with the authors' permission. *Brickify* exemplifies how a single tool can manifest diverse levels of support across all three framework dimensions. Numbered markers map interface elements to these dimensions, with support levels following a 0–3 scale encoded by fill style: darker-filled boxes indicate strong support (3), light-filled boxes moderate support (2), white boxes limited support (1), and grey boxes no support (0). The tool demonstrates strong support for global and component control, direct manipulation, iterative adjustments, ideation, and editing; moderate support for attribute-based input, GUI interactions, exploration, blending, variation exploration, and prompt guidance; and limited or no support for parameter control, precision control, or evaluation. Scores reflect both presence and depth of functionality (e.g., manipulable attribute tokens without continuous sliders are rated 2, while component resizing without fine-grained precision is rated 1). Additional contrasting examples, illustrating representative variations across each main framework dimension, are presented in Appendix A.5.

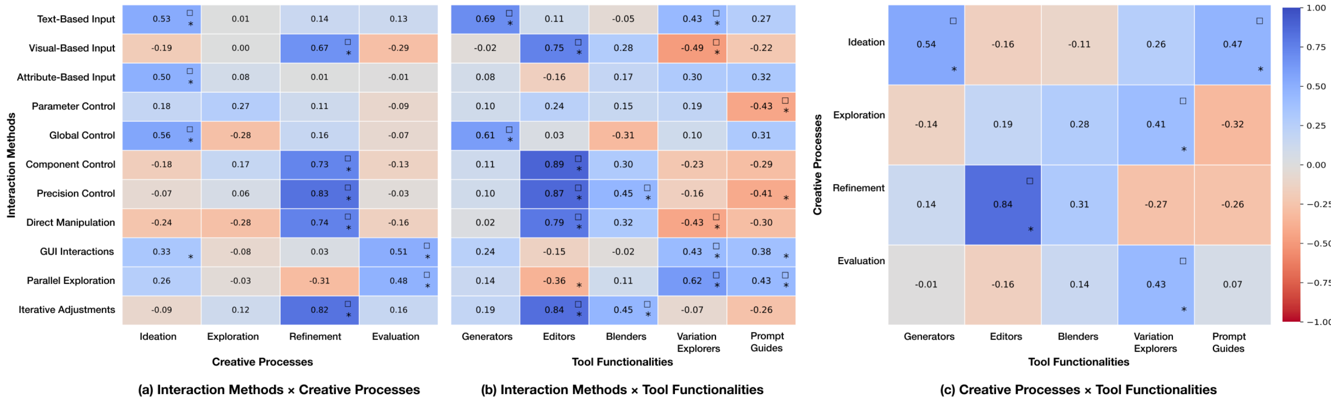


Figure 5: Cross-framework correlations across 37 tools. (a) Interaction Methods and Creative Processes, (b) Interaction Methods and Tool Functionalities, and (c) Creative Processes and Tool Functionalities. Markers indicate statistical significance: a star (∗) in the bottom-right of a cell denotes $p < .05$, and a square (□) in the top-right of a cell denotes FDR-corrected $q < .05$. Correlation values range from −1 (perfect negative) to +1 (perfect positive). The heatmaps highlight consistent alignments (e.g., *Text-Based Input* with *Generators* and *Ideation*, *Visual-Based Input* with *Editors*) as well as emerging patterns (e.g., weaker links between *Variation Explorers* and *Visual-Based Input*, and between *Prompt Guides* and fine-grained controls such as *Parameter Control* or *Precision Control*).

text-based input, ideation, and global control, while visual-based input, parameter control, precision control, refinement, and evaluation receive comparatively weaker support. Building on these empirical patterns, this section explains how each design opportunity responds to the identified structural tendencies and how future interfaces can better support creative workflows. Each subsection concludes with a **Principle** (derived from RQ3 patterns) and one or more **Design Opportunities (DOs)**.

5.1 Visual-Based Inputs: Opportunities for Expansion

Table 1 shows that only 13/37 tools provide strong support for visual-based input. Figure 5 further indicates that visual modalities correlate with *Refinement* and *Editor* functionalities, but not with *Ideation*, *Exploration*, or *Generators*. This positions visual interaction mostly as a late-stage refinement tool rather than a capability integrated across creative phases. While tools such as Firefly, Vizcom, and Inkspire [55] demonstrate the potential of visually driven workflows, most systems tie visual input closely to text prompting, limiting support for visually grounded practices [79].

DO1: Enhancing Visual-Based Interaction. **Principle.** Strong *visual-based input* is present in only a subset of tools (13/37), and typically offers limited expressive control. **Opportunity.** Extend visual input beyond reference images to support more expressive interactions such as *sketching*, *color blocking*, and *visual attribute pickers* (e.g., Brickify's tokens [99], FusAIn's pen-applied attributes [83], Inkspire's sketch-to-image loops [55]). These approaches allow users to directly express spatial structure and visual intent, which is especially beneficial for mid- to late-stage tasks requiring spatial specificity and fine-grained control [78, 99], while being less critical during early ideation, which is dominated by text and global controls [50].

DO2: Bridging Visual Input and Variation Exploration. **Principle.** *Variation explorers* are negatively correlated with *visual input.* **Opportunity.** Because tools rarely combine visual input (e.g., sketching, reference images) with structured variation exploration, interfaces should more tightly integrate the two. For example, users could upload or sketch an initial input and then explore variations through grids, clusters, or branching trees, with smooth transitions between input and exploration views. As shown in Figure 5, the negative correlation between *Variation Explorers* and visual input underscores the need for such integration.

5.2 Strong Focus on Semantic Attributes, Limited Parameter Control

Table 1 indicates that parameter control is one of the least supported interaction types, with most tools exposing semantic attributes rather than inference-level parameters. Midjourney remains the only tool offering broadly comprehensive parameter controls. Five tools provide moderate support (e.g., batch size, prompt weighting, and seed), and most academic systems offer limited or none. Midjourney's command-based interface enables fine-grained tuning but can feel complex to novices [79, 100, 102, 110]. Demand for more accessible parameter control is evident in user communities [20, 90] and in wrapper tools such as PromptFolder [85], which translate natural language into Midjourney-compatible commands. This trend raises an open question: **How much parameter control do users need, and how can interfaces adapt to different levels of expertise?** Current tools take varied approaches: DreamStudio and RunwayML hide advanced menus that users can expand on demand, Stable Diffusion WebUI [4] exposes extensive customization options, and PromptPaint [18] offers a visual palette as an alternative to numeric input.

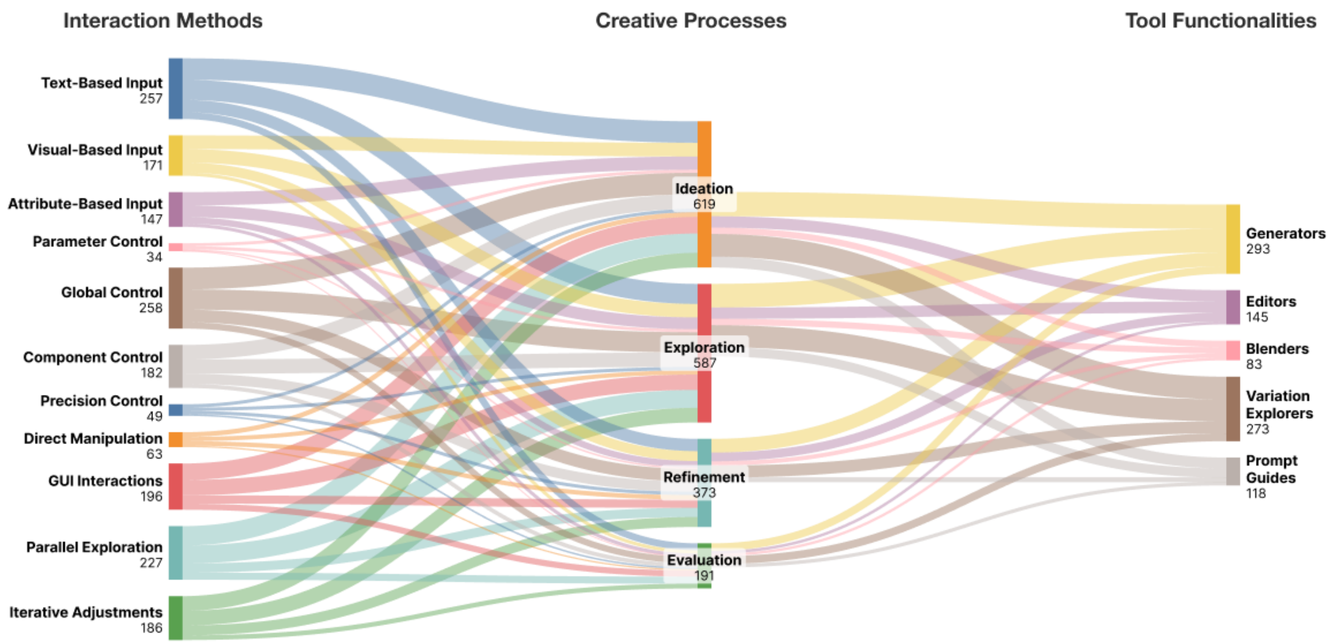


Figure 6: Global overview of tool distributions across frameworks. The Sankey diagram summarizes co-support across *Interaction Methods*, *Creative Processes*, and *Tool Functionalities*. Flows are computed using an ordinal co-support scoring model (strong–strong = 3, strong–moderate = 2, moderate–moderate = 1, otherwise = 0), with these scores determining flow widths. This approach avoids interval-scale assumptions, suppresses weak or incidental links, and highlights dominant pathways and underrepresented patterns in the multi-dimensional design space.

DO3: Simplifying Parameter Control. **Principle.** *Ideation* co-occurs with *text-based input* and *global control*, while *prompt guides* seldom co-occur with *parameter control* and *precision control*. **Opportunity.** Provide more accessible mechanisms for adjusting both parameters and semantic attributes. Frequently used settings can be supported through sliders or visual metaphors (e.g., PromptPaint's palette [18]), while advanced panels remain collapsible to prevent overwhelming novices and still accommodate expert use. The negative correlation between parameter control and prompt guidance in Figure 5 suggests the value of designs that reconcile these functions.

5.3 Lack of Precision and Direct Control

Table 1 shows that precision control is rare (only three tools scored strong support), indicating a structural gap. Existing interfaces emphasize global or component-level editing but offer limited fine-grained control. Users requiring spatial accuracy frequently supplement GenAI tools with traditional editors [79]. Tools such as ProtoDreamer's anchor-based resizing, Firefly's region editing, and Brickify's token manipulation explore direct control, while prior work shows sketch input can outperform text for refinement [50].

DO4: Enabling Precision Editing. **Principle.** *Refinement* aligns with *editors*, *precision control*, and *iterative adjustments*. **Opportunity.** Incorporate features such as region-aware selection, sketch-based refinement, anchor-point resizing, drag-and-drop manipulation, layering, and inpainting to better support precision-oriented workflows and lessen dependence on external design tools. These capabilities may be particularly useful in domains where spatial accuracy and local adjustments matter (e.g., product, interior, game, or interface design), though needs differ across practices and user groups.

DO5: Integrating Parallel Exploration with Direct Manipulation. **Principle.** *Parallel Exploration* is negatively correlated with *Direct Manipulation*. **Opportunity.** Structured comparison and hands-on editing are rarely integrated. Interfaces should allow fluid transitions between exploration views (e.g., grids, clusters) and direct manipulation while preserving design history (such as prompts, seeds, and operations) across states. Figure 5 shows a negative association between these modes, indicating an opportunity to enable users to explore options and refine them through direct manipulation within the same workspace.

5.4 Strong Support for Ideation and Exploration, Gaps in Refinement and Evaluation

Table 2 shows strong support for ideation and exploration but limited support for refinement and evaluation, indicating that current tools prioritize early-stage breadth. Systems occasionally integrate comparison utilities (e.g., heatmaps, similarity matrices) [14, 133], but structured evaluation remains rare, and moodboard-style environments rely heavily on manual inspection.

DO6: Integrating Evaluation, Exploration, and Editing Workflows. **Principle.** *Evaluation* benefits from structured graphical interfaces and *parallel exploration*, yet these features are rarely integrated with *editors.* **Opportunity.** Provide dedicated evaluation views (e.g., side-by-side grids, rating panels, criteria tags) that leverage *parallel exploration* for comparison and selection, while enabling smooth transitions to editing (e.g., inpainting, layering) and preserving provenance such as prompts, seeds, and edit histories. Figure 5 further shows that evaluation aligns with exploration but not editing, indicating an opportunity to more tightly connect these workflows.

DO7: Enabling Defaults and Rapid Re-Prompting. **Principle.** The ideation process benefits from *global controls* and fast loops. **Opportunity.** Make rapid exploration the default. Provide one-click presets (e.g., style, tone, composition), re-prompt shortcuts (e.g., "more/less like this"), and history-based presets that adapt to a user's prior prompts or design choices. Use simple sliders or toggles for key constraints such as style or aspect ratio. These features enable users to quickly explore creative directions before engaging with more detailed controls, aligning with patterns in Figure 5 showing ideation's reliance on text and global controls rather than fine-grained parameters.

5.5 Expanding GUI Interactions

Table 1 shows broad but shallow GUI support, indicating room for more adaptive designs. Many systems rely on basic sliders, buttons, and dropdowns for adjusting attributes [17, 82, 124]. More advanced interfaces include graph and tree views for arranging components or tracking variations [22, 38]. Commercial tools remain text-centric, though some now introduce sliders for refinement [67]. Node-based systems such as ComfyUI support more complex workflows (e.g., blending, parameter tuning, provenance tracking), providing richer control for refinement and evaluation [21].

DO8: Designing Flexible GUI Layers. **Principle.** Creative workflows span quick adjustments and complex refinement, and users differ widely in expertise. **Opportunity.** Provide tiered GUIs: simple controls such as guided sliders and drag-and-drop for novices, and node- or layer-based interfaces for advanced users. Keep complex panels collapsible to make advanced functionality more approachable and ease the learning curve for novices. Interfaces may also recall prior configurations so users can reopen or extend earlier workflows. Figure 5 shows GUI interactions correlate with evaluation, suggesting opportunities to extend GUI support into refinement tasks.

5.6 Prompt Guidance: Academic Tools Ahead of Commercial

Table 3 shows strong prompt-guidance support in academic systems but limited integration in commercial ones. Academic tools [17, 103, 124] provide strong prompt guidance through keyword extraction, contextual suggestions, and interactive support. Midjourney's `describe` feature can extract attributes such as subject, artist, and style from uploaded images, but it operates separately from the main image-generation workflow. Despite widespread recognition that *text prompting* is difficult, guidance in commercial systems remains limited compared to academic advances. Promising directions include real-time feedback, interactive scaffolds, and partially more automated prompting workflows. This raises a broader question: **Should GenAI systems focus on improving *text prompting*, or explore alternative interaction paradigms that move beyond text?**

DO9: Enhancing Prompt Assistance. **Principle.** *Prompt guidance* supports *ideation* but remains under-integrated in commercial tools. **Opportunity.** Integrate prompt guidance throughout the generation workflow. Provide keyword suggestions, example prompts, and editable scaffolds with brief explanations of prompt effects. Extend guidance into later stages with edit-oriented prompts (e.g., inpainting, adjustments), explanations for differences between variants, and suggestions for next steps in *evaluation* or *refinement.* Guidance can also adapt to users' prompt histories by surfacing recurring styles or terms. Figure 5 shows prompt guidance aligns with ideation, pointing to opportunities for extension into later phases.

6 Limitations

This study has several limitations. First, our review focuses on tools reported in papers from HCI venues sponsored by SIGCHI, accessed through the ACM DL, with additional pre-screening of selected non-HCI venues. While this ensured relevance to interaction design, it may have excluded tools published in other databases or disciplinary communities. Second, our analysis centers on user interaction and interface design rather than technical or model-level aspects. As a result, we do not examine model architectures, training pipelines, or how improvements in underlying generative models may influence interface behavior. Third, our selection of commercial tools focused on widely used platforms in design practice, potentially overlooking emerging systems with smaller user bases or limited documentation. A further limitation concerns our scope definition. Restricting the review to the January 2022–July 2025 period and to 2D, screen-based GenAI tools shapes the interaction patterns we observe. Earlier systems identified in our contrast set (Appendix A.1) were oriented toward model-guided sketch transformations or narrow-scope outputs rather than open-ended, user-facing image generation. More broadly, some single-category generators (e.g., face-specific systems) fall outside our inclusion criteria because they do not generalize to open-ended creative workflows. Spatial modalities (VR/AR/3D) rely on qualitatively different interaction paradigms, such as embodied manipulation, viewpoint-dependent inspection, and in-environment spatial arrangement. Including such systems would require extending our criteria to cover 3D object, asset, or scene-generation workflows. While these modalities offer promising directions for expanding the framework, incorporating them here would broaden and potentially diffuse the analytical focus of this 2D review. We also acknowledge that excluding theoretical and qualitative-only studies may omit conceptual insights, although this decision supports consistent comparison across implemented systems. Finally, our evaluation relied on a scoring-based classification of support levels (strong, moderate, limited, none) within each framework category. Although criteria were applied rigorously and iteratively, interpretation differences and variation in available tool documentation may have influenced some judgments. Future work could address these limitations by drawing from broader disciplinary sources, exploring underrepresented tools, and diversifying research methods. Future studies may also incorporate spatial or non-HCI systems and employ broader evaluation methods (e.g., expert panels or user studies) to validate and extend the framework across diverse modalities and contexts.

7 Conclusion and Future Work

This work presents a systematic review of 37 GenAI image tools, comprising 28 from HCI research and nine from industry, with a focus on interaction design and interface functionality. We introduced a design space structured around three analytical frameworks—interaction methods, creative processes, and tool functionalities. Using a scoring system within these frameworks, combined with correlation analysis, we identified gaps, recurring patterns, and design opportunities in current GenAI image tools. Text-based prompts remain the dominant input modality, while visual- and attribute-based inputs are increasingly integrated, particularly in academic systems. These inputs typically complement rather than replace text, especially during refinement. Our analysis also revealed limited support for precision control and parameter control, with most tools prioritizing global or component-level adjustments over fine-grained editing. Advanced GUI interactions and structured evaluation workflows remain underdeveloped in current tools, despite their importance for professional design practice. While ideation and exploration are strongly supported, refinement, detailed editing, and prompt guidance remain comparatively weak.

From these findings, we surfaced nine design opportunities spanning: (1) input modalities (enhancing visual interaction, linking visual input with variation exploration), (2) control mechanisms (simplifying parameter control, enabling precision editing), (3) workflow integration (connecting exploration with direct manipulation, integrating evaluation with exploration and editing), and (4) interface scaffolds (defaults and rapid re-prompting, flexible GUI layers, prompt assistance for later stages). Together, these opportunities highlight directions for more usable, adaptable, and creativity-supportive GenAI image tool interfaces. Future work should investigate how these insights can inform new interaction techniques, UI components, and prototyping toolkits for GenAI-enabled creative workflows. As the field evolves, we plan to extend this review by incorporating newer tools and user-centered evaluations to further refine and validate the design space. Future research may also broaden the scope beyond HCI venues and 2D screen-based tools, incorporating spatial interfaces and non-HCI systems to examine how interaction patterns vary across modalities and disciplinary contexts. Building on our analytical frameworks, this work provides a foundation for future HCI research on GenAI image tool interfaces that more effectively support user needs throughout the creative process.

Acknowledgments

We thank the anonymous reviewers for their constructive feedback. Generative AI tools were used to improve readability (e.g., rephrasing and formatting). The authors developed all substantive content, analyses, and conclusions and take full responsibility for the work.

References

[1] 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
[2] Adobe. 2025. Adobe Firefly. https://www.adobe.com/products/firefly.html.
[3] Shm Garanganao Almeda, J.D. 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*. ACM, Honolulu HI USA, 1–17. https://doi.org/10.1145/3613904.3642858
[4] AUTOMATIC1111. 2025. Stable Diffusion Web UI. https://github.com/AUTOMATIC1111/stable-diffusion-webui
[5] Majid Behravan, Maryam Haghani, and Denis Gracanin. 2025. Transcending Dimensions using Generative AI: Real-Time 3D Model Generation in Augmented Reality. https://doi.org/10.48550/arXiv.2504.21033
[6] Volker Bilgram and Felix Laarmann. 2023. Accelerating Innovation With Generative AI: AI-Augmented Digital Prototyping and Innovation Methods. *IEEE Engineering Management Review* 51, 2 (June 2023), 18–25. https://doi.org/10.1109/EMR.2023.3272799
[7] Doug Bowman, Jian Chen, Chadwick Wingrave, John Lucas, Andrew Ray, Nicholas Polys, Qing Li, Yonca Haciahmetoglu, Seonho Kim, Robert Boehringer, and Tao Ni. 2006. New Directions in 3D User Interfaces. *IJVR* 5 (Jan. 2006), 3–14. https://doi.org/10.20870/IJVR.2006.5.2.2683
[8] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. 2004. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., USA.
[9] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3586183.3606725
[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa
[11] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.*, Harris Cooper, Paul M. Camic, Debra L. Long, A. T. Panter, David Rindskopf, and Kenneth J. Sher (Eds.). American Psychological Association, Washington, 57–71. https://doi.org/10.1037/13620-004
[12] Rina Caballar. 2024. Generative AI vs. Predictive AI: What's the Difference? https://www.ibm.com/blog/generative-ai-vs-predictive-ai-whats-the-difference.
[13] Minsuk Chang, Stefania Druga, Alexander J. Fiannaca, Pedro Vergani, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. The Prompt Artists. In *Proceedings of the 15th Conference on Creativity and Cognition (C&C '23)*. Association for Computing Machinery, New York, NY, USA, 75–87. https://doi.org/10.1145/3591196.3593515
[14] Liuqing Chen, Qianzhi Jing, Yixin Tsang, Qianyi Wang, Ruocong Liu, Duowei Xia, Yunzhan Zhou, and Lingyun Sun. 2024. AutoSpark: Supporting Automobile Appearance Design Ideation with Kansei Engineering and Generative AI. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, Pittsburgh PA USA, 1–19. https://doi.org/10.1145/3654777.3676337
[15] Xiang Anthony Chen, Tiffany Knearem, and Yang Li. 2025. The GenUI Study: Exploring the Design of Generative UI Tools to Support UX Practitioners and Beyond. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 1179–1196. https://doi.org/10.1145/3715336.3735780
[16] Li-Yuan Chiou, Peng-Kai Hung, Rung-Huei Liang, and Chun-Teng Wang. 2023. Designing with AI: An Exploration of Co-Ideation with Image Generators. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 1941–1954. https://doi.org/10.1145/3563657.3596001
[17] DaEun Choi, Sumin Hong, Jeongeon Park, John Joon Young Chung, and Juho Kim. 2024. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–25. https://doi.org/10.1145/3613904.3642794
[18] John Joon Young Chung and Eytan Adar. 2023. PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–17. https://doi.org/10.1145/3586183.3606777

[19] John Joon Young Chung, Shiqing He, and Eytan Adar. 2021. The Intersection of Users, Roles, Interactions, and Technologies in Creativity Support Tools. In *Designing Interactive Systems Conference 2021*. ACM, Virtual Event USA, 1817–1833. https://doi.org/10.1145/3461778.3462050
[20] Civitai. 2025. Civitai. https://civitai.com.
[21] ComfyUI. 2025. ComfyUI. https://www.comfy.org.
[22] Hai Dang, Frederik Brudy, George Fitzmaurice, and Fraser Anderson. 2023. WorldSmith: Iterative and Expressive Prompting for World Building with a Generative AI. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3586183.3606772
[23] Richard Lee Davis, Thiemo Wambsganss, Wei Jiang, Kevin Gonyop Kim, Tanja Käser, and Pierre Dillenbourg. 2024. Fashioning Creative Expertise with Generative AI: Graphical Interfaces for Design Space Exploration Better Support Ideation Than Text Prompts. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–26. https://doi.org/10.1145/3613904.3642908
[24] Ben Dickson. 2022. How 2022 Became the Year of Generative AI. https://venturebeat.com/ai/how-2022-became-the-year-of-generative-ai. Venturebeat.
[25] DreamStudio. 2025. DreamStudio. https://dreamstudio.ai.
[26] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch, and Patrick Zschech. 2024. Generative AI. *Business & Information Systems Engineering* 66, 1 (Feb. 2024), 111–126. https://doi.org/10.1007/s12599-023-00834-7
[27] Tham Yik Foong, Shashank Kotyan, Po Yuan Mao, and Danilo Vasconcellos Vargas. 2023. The Challenges of Image Generation Models in Generating Multi-Component Images. https://doi.org/10.48550/arXiv.2311.13620
[28] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–18. https://doi.org/10.1145/3290605.3300619
[29] Mengyu Gao and Qiulei Dong. 2024. Adaptive Conditional Denoising Diffusion Model With Hybrid Affinity Regularizer for Generalized Zero-Shot Learning. *IEEE Transactions on Circuits and Systems for Video Technology* 34, 7 (July 2024), 5641–5652. https://doi.org/10.1109/TCSVT.2024.3359238
[30] Frederic Gmeiner, Humphrey Yang, Lining Yao, Kenneth Holstein, and Nikolas Martelaro. 2023. Exploring Challenges and Opportunities to Support Designers in Learning to Co-create with AI-based Manufacturing Design Tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–20. https://doi.org/10.1145/3544548.3580999
[31] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Networks. https://arxiv.org/abs/1406.2661
[32] Shunan Guo, Zhuochen Jin, Fuling Sun, Jingwen Li, Zhaorui Li, Yang Shi, and Nan Cao. 2021. Vinci: An Intelligent Graphic Design System for Generating Advertising Posters. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–17. https://doi.org/10.1145/3411764.3445117
[33] Shihui Guo, Yubin Shi, Pintong Xiao, Yinan Fu, Juncong Lin, Wei Zeng, and Tong-Yee Lee. 2023. Creative and Progressive Interior Color Design with Eye-tracked User Preference. *ACM Transactions on Computer-Human Interaction* 30, 1 (Feb. 2023), 1–31. https://doi.org/10.1145/3542922
[34] Yuhan Guo, Hanning Shao, Can Liu, Kai Xu, and Xiaoru Yuan. 2025. PrompTHis: Visualizing the Process and Influence of Prompt Editing during Text-to-Image Creation. *IEEE Transactions on Visualization and Computer Graphics* 31, 9 (2025), 4547–4559. https://doi.org/10.1109/TVCG.2024.3408255
[35] Yaru Hao, Zewen Chi, Li Dong, and Furu Wei. 2023. Optimizing Prompts for Text-to-Image Generation. In *Proceedings of the 37th International Conference on Neural Information Processing Systems (NeurIPS '23, Vol. 36)*. Curran Associates, Inc., Red Hook, NY, USA, 66923–66939. https://proceedings.neurips.cc/paper_files/paper/2023/file/d346d91999074dd8d6073d4c3b13733b-Paper-Conference.pdf
[36] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. https://doi.org/10.48550/arXiv.2006.11239
[37] Bofei Huang and Haoran Xie. 2025. PromptNavi: Text-to-image generation through interactive prompt visual exploration. *Computers & Graphics* 132 (Nov. 2025), 104417. https://doi.org/10.1016/j.cag.2025.104417
[38] 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 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–19. https://doi.org/10.1145/3613904.3642824
[39] Hugging Face. 2024. Stable Diffusion Documentation. https://huggingface.co/docs/diffusers/en/stable_diffusion.
[40] Angel Hsing-Chi Hwang. 2022. Too Late to be Creative? AI-Empowered Tools in Creative Processes. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–9. https://doi.org/10.1145/3491101.3503549
[41] Nanna Inie, Jeanette Falk, and Steve Tanimoto. 2023. Designing Participatory AI: Creative Professionals' Worries and Expectations about Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–8. https://doi.org/10.1145/3544549.3585657
[42] 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*. ACM, Yokohama Japan, 1–18. https://doi.org/10.1145/3411764.3445093
[43] Everypixel Journal. 2023. AI Image Statistics for 2024: How Much Content Was Created by AI. https://journal.everypixel.com/ai-image-statistics
[44] Pegah Karimi, Jeba Rezwana, Safat Siddiqui, Mary Lou Maher, and Nasrin Dehbozorgi. 2020. Creative sketching partner: an analysis of human-AI co-creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. ACM, Cagliari Italy, 221–230. https://doi.org/10.1145/3377325.3377522
[45] Diederik P. Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013). https://indico.math.cnrs.fr/event/11377/attachments/4589/6915/18012024_Kingma-and-Welling-2022%20Auto-Encoding%20Variational%20Bayes.pdf
[46] Will Knight. 2023. Where the AI Art Boom Came From—and Where It's Going. https://www.wired.com/gallery/where-the-ai-art-boom-came-from-and-where-its-going/ WIRED.
[47] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 6 (May 2017), 84–90. https://doi.org/10.1145/3065386
[48] Chinmay Kulkarni, Stefania Druga, Minsuk Chang, Alex Fiannaca, Carrie Cai, and Michael Terry. 2023. A Word is Worth a Thousand Pictures: Prompts as AI Design Material. https://doi.org/10.48550/arXiv.2303.12647
[49] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. https://doi.org/10.2307/2529310 Publisher: International Biometric Society.
[50] Seung Won Lee, Tae Hee Jo, Semin Jin, Jiin Choi, Kyungwon Yun, Sergio Bromberg, Seonghoon Ban, and Kyung Hoon Hyun. 2024. The Impact of Sketch-guided vs. Prompt-guided 3D Generative AIs on the Design Exploration Process. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. https://doi.org/10.1145/3613904.3642218
[51] Joanne Leong, David Ledo, Thomas Driscoll, Tovi Grossman, George Fitzmaurice, and Fraser Anderson. 2025. Paratrouper: Exploratory Creation of Character Cast Visuals Using Generative AI. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3706598.3714242
[52] Lexica. 2024. Lexica. https://lexica.art.
[53] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. https://doi.org/10.1145/3613904.3642114
[54] Q. Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly Understanding: Information Needs for Model Transparency to Support Design Ideation for AI-Powered User Experience. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3544548.3580652
[55] 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. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3706598.3713397
[56] Haichuan Lin, Yilin Ye, Jiazhi Xia, and Wei Zeng. 2025. SketchFlex: Facilitating Spatial-Semantic Coherence in Text-to-Image Generation with Region-Based Sketches. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3706598.3713801
[57] Vivian Liu. 2023. Beyond Text-to-Image: Multimodal Prompts to Explore Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–6. https://doi.org/10.1145/3544549.3577043
[58] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–23. https://doi.org/10.1145/3491102.3501825
[59] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3526113.3545621
[60] lllyasviel. 2023. ControlNet. https://github.com/lllyasviel/ControlNet.

[61] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully Convolutional Networks for Semantic Segmentation. https://doi.org/10.48550/arXiv.1411.4038

[62] Reuben Luera, Ryan A. Rossi, Alexa Siu, Franck Dernoncourt, Tong Yu, Sungchul Kim, Ruiyi Zhang, Xiang Chen, Hanieh Salehy, Jian Zhao, Samyadeep Basu, Puneet Mathur, and Nedim Lipka. 2024. Survey of User Interface Design and Interaction Techniques in Generative AI Applications. https://doi.org/10.48550/arXiv.2410.22370

[63] Atefeh Mahdavi Goloujeh, Anne Sullivan, and Brian Magerko. 2024. Is It AI or Is It Me? Understanding Users' Prompt Journey with Text-to-Image Generative AI Tools. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3613904.3642861

[64] Mango. 2024. Mango creates the first campaign generated by artificial intelligence for its Teen line. https://www.mangofashiongroup.com/en/w/mango-crea-la-primera-campa%C3%B1a-generada-con-inteligencia-artificial-para-su-l%C3%ADnea-teen

[65] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. DirectGPT: A Direct Manipulation Interface to Interact with Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3613904.3642462

[66] Mario Michelessa, Jamie Ng, Christophe Hurter, and Brian Y. Lim. 2025. Varif.ai to Vary and Verify User-Driven Diversity in Scalable Image Generation. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 1867–1885. https://doi.org/10.1145/3715336.3735847

[67] Midjourney. 2025. Midjourney. https://www.midjourney.com

[68] David Moher, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Journal of Clinical Epidemiology* 62, 10 (Oct. 2009), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005 Publisher: Elsevier.

[69] Michael Muller, Lydia B Chilton, Anna Kantosalo, Charles Patrick Martin, and Greg Walsh. 2022. GenAICHI: Generative AI and HCI. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–7. https://doi.org/10.1145/3491101.3503719

[70] NightCafe. 2025. NightCafe. https://creator.nightcafe.studio/.

[71] Nike. 2024. Nike Unveils AI-Driven A.I.R. Prototype Sneakers. https://www.soleretriever.com/news/articles/nike-ai-driven-air-prototype-sneakers-13-athletes-detailed-look-april-2024.

[72] Notion. 2025. Notion. https://www.notion.com.

[73] OpenAI. 2024. Editing Images with DALL·E. https://help.openai.com/en/articles/9055440-editing-your-images-with-dall-e.

[74] OpenAI. 2025. DALL·E 3. https://openai.com/index/dall-e-3/.

[75] OpenAI. 2025. DALL·E Outpainting. https://openai.com/index/dall-e-introducing-outpainting/.

[76] Kostas Ordoumpozanis, Markos Konstantakis, Stavroula Zoi, and George Caridakis. 2025. Generative AI: A Systematic Review of Related Interfaces and Interactions. In *Proceedings of the 3rd International Conference of the ACM Greek SIGCHI Chapter (CHIGreece '25)*. Association for Computing Machinery, New York, NY, USA, 39–47. https://doi.org/10.1145/3749012.3749052

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

[78] Hyerim Park, Malin Eiband, Andre Luckow, and Michael Sedlmair. 2025. Exploring Visual Prompts: Refining Images with Scribbles and Annotations in Generative AI Image Tools. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3706599.3719802

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

[80] Hyerim Park, Joscha Eirich, Andre Luckow, and Michael Sedlmair. 2025. Design and Evaluation of a Generative AI-Driven VR Texturing Tool: A Design Science Approach. In *Proceedings of the European Conference on Information Systems (ECIS 2025)*. Association for Information Systems, Amman, Jordan, 16 pages. https://aisel.aisnet.org/ecis2025/hci/hci/2

[81] Nikita Pavlichenko and Dmitry Ustalov. 2023. Best Prompts for Text-to-Image Models and How to Find Them. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Taipei Taiwan, 2067–2071. https://doi.org/10.1145/3539618.3592000

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

[83] Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2025. FusAIn: Composing Generative AI Visual Prompts Using Pen-based Interaction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3706598.3714027

[84] Savvas Petridis, Michael Terry, and Carrie J. Cai. 2023. PromptInfuser: How Tightly Coupling AI and UI Design Impacts Designers' Workflows. https://doi.org/10.48550/arXiv.2310.15435

[85] Promptfolder. 2025. Midjourney Prompt Generator. https://promptfolder.com/midjourney-prompt-helper/.

[86] PromptHero. 2025. PromptHero. https://prompthero.com/ PromptHero.

[87] Han Qiao, Vivian Liu, and Lydia Chilton. 2022. Initial Images: Using Image Prompts to Improve Subject Representation in Multimodal AI Generated Art. In *Creativity and Cognition*. ACM, Venice Italy, 15–28. https://doi.org/10.1145/3527927.3532792

[88] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. https://doi.org/10.48550/arXiv.2204.06125

[89] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-Shot Text-to-Image Generation. https://doi.org/10.48550/arXiv.2102.12092

[90] Reddit. 2025. r/midjourney. https://www.reddit.com/r/midjourney/.

[91] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems*, Vol. 28. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf

[92] Bernhard E. Riecke, Joseph J. LaViola, and Ernst Kruijff. 2018. 3D user interfaces for virtual reality and games: 3D selection, manipulation, and spatial navigation. In *ACM SIGGRAPH 2018 Courses (SIGGRAPH '18)*. Association for Computing Machinery, New York, NY, USA, 1–94. https://doi.org/10.1145/3214834.3214869

[93] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042 ISSN: 2575-7075.

[94] Runway. 2025. Runway. https://runwayml.com.

[95] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J. Fleet, and Mohammad Norouzi. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. https://doi.org/10.48550/arXiv.2205.11487

[96] Iqbal H. Sarker. 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. *Sn Computer Science* 2, 3 (2021), 160. https://doi.org/10.1007/s42979-021-00592-x

[97] Yulin Shen, Yifei Shen, Jiawen Cheng, Chutian Jiang, Mingming Fan, and Zeyu Wang. 2024. Neural Canvas: Supporting Scenic Design Prototyping by Integrating 3D Sketching and Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. https://doi.org/10.1145/3613904.3642096

[98] Jingyu Shi, Rahul Jain, Hyungjun Doh, Ryo Suzuki, and Karthik Ramani. 2024. An HCI-Centric Survey and Taxonomy of Human-Generative-AI Interactions. https://doi.org/10.48550/arXiv.2310.07127

[99] Xinyu Shi, Yinghou Wang, Ryan Rossi, and Jian Zhao. 2025. Brickify: Enabling Expressive Design Intent Specification through Direct Manipulation on Design Tokens. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–20. https://doi.org/10.1145/3706598.3714087

[100] Ben Shneiderman. 1983. Direct Manipulation: A Step Beyond Programming Languages. *Computer* 16, 8 (Aug. 1983), 57–69. https://doi.org/10.1109/MC.1983.1654471

[101] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (Dec. 2007), 20–32. https://doi.org/10.1145/1323688.1323689

[102] Ben Shneiderman and Pattie Maes. 1997. Direct manipulation vs. interface agents. *Interactions* 4, 6 (Nov. 1997), 42–61. https://doi.org/10.1145/267505.267514

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

[104] Zineb Sordo, Eric Chagnon, and Daniela Ushizima. 2025. A Review on Generative AI For Text-To-Image and Image-To-Image Generation and Implications To Scientific Images. https://doi.org/10.48550/arXiv.2502.21151

[105] StarryAI. 2025. StarryAI. https://starryai.com.

[106] Statista. 2025. Global generative AI image tool market share 2023. https://www.statista.com/forecasts/1423989/world-generative-ai-image-tool-market-share Statista.

[107] Zhida Sun, Zhenyao Zhang, Yue Zhang, Min Lu, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. 2025. Creative Blends of Visual Concepts. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3706598.3713683

[108] Supriyono, Aji Prasetya Wibawa, Suyono, and Fachrul Kurniawan. 2024. Advancements in natural language processing: Implications, challenges, and future directions. *Telematics and Informatics Reports* 16 (Dec. 2024), 100173. https://doi.org/10.1016/j.teler.2024.100173

[109] Kamal Taha, Paul D. Yoo, Chan Yeun, Dirar Homouz, and Aya Taha. 2024. A comprehensive survey of text classification techniques and their research applications: Observational and experimental insights. *Computer Science Review* 54 (Nov. 2024), 100664. https://doi.org/10.1016/j.cosrev.2024.100664

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

[111] Sirui Tao, Ivan Liang, Cindy Peng, Zhiqing Wang, Srishti Palani, and Steven P. Dow. 2025. DesignWeaver: Dimensional Scaffolding for Text-to-Image Product Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–26. https://doi.org/10.1145/3706598.3714211

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

[113] Rob Toews. 2022. 4 Predictions About The Wild New World Of Text-To-Image AI. https://www.forbes.com/sites/robtoews/2022/09/11/4-hot-takes-about-the-wild-new-world-of-generative-ai/ Section: AI.

[114] Severi Uusitalo, Antti Salovaara, Tero Jokela, and Marja Salmimaa. 2024. "Clay to Play With": Generative AI Tools in UX and Industrial Design Practice. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 1566–1578. https://doi.org/10.1145/3643834.3661624

[115] Cyrus Vachha, Yixiao Kang, Zach Dive, Ashwat Chidambaram, Anik Gupta, Eunice Jun, and Björn Hartmann. 2025. Dreamcrafter: Immersive Editing of 3D Radiance Fields Through Flexible, Generative Inputs and Outputs. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3706598.3714312

[116] Mathias Peter Verheijden and Mathias Funk. 2023. Collaborative Diffusion: Boosting Designerly Co-Creation with Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–8. https://doi.org/10.1145/3544549.3585680

[117] Visual Electric. 2025. Visual Electric. https://visualelectric.com.

[118] Vizcom. 2025. Vizcom. https://www.vizcom.ai.

[119] Qian Wan and Zhicong Lu. 2023. GANCollage: A GAN-Driven Digital Mood Board to Facilitate Ideation in Creativity Support. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 136–146. https://doi.org/10.1145/3563657.3596072

[120] Nan Wang, Hyunsuk Kim, Junfeng Peng, and Jiayi Wang. 2025. Exploring creativity in human–AI co-creation: a comparative study across design experience. *Frontiers in Computer Science* 7 (Sept. 2025). https://doi.org/10.3389/fcomp.2025.1672735 Publisher: Frontiers.

[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*. ACM, Honolulu HI USA, 1–20. https://doi.org/10.1145/3613904.3642901

[122] Wen-Fan Wang, Chien-Ting Lu, Nil Ponsa i Campanyà, Bing-Yu Chen, and Mike Y. Chen. 2025. AIdeation: Designing a Human-AI Collaborative Ideation System for Concept Designers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–28. https://doi.org/10.1145/3706598.3714148

[123] Ye Wang, Nicole B. Damen, Thomas Gale, Voho Seo, and Hooman Shayani. 2024. Inspired by AI? A Novel Generative AI System To Assist Conceptual Automotive Design. https://doi.org/10.48550/arXiv.2407.11991

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

[125] Christoph Johannes Weber, Jenny Huang, and Sylvia Rothe. 2025. Drawing-in-Steps: Supporting Creative Goals through User Engagement via Hierarchical Image Generation. In *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences (IMX '25)*. Association for Computing Machinery, New York, NY, USA, 108–125. https://doi.org/10.1145/3706370.3727862

[126] Di Wu, Mingzhu Wu, Yeye Li, Jianan Jiang, Xinglin Li, Hanhui Deng, Can Liu, and Yi Li. 2024. StyleWe: Towards Style Fusion in Generative Fashion Design with Efficient Federated AI. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2 (Nov. 2024), 515:1–515:31. https://doi.org/10.1145/3687054

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

[128] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J Cai. 2022. PromptChainer: Chaining Large Language Model Prompts through Visual Programming. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3491101.3519729

[129] Zhenqiang Ying and Alan Bovik. 2020. 180-degree Outpainting from a Single Image. https://doi.org/10.48550/arXiv.2001.04568

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

[131] Chengzhi Zhang, Weijie Wang, Paul Pangaro, Nikolas Martelaro, and Daragh Byrne. 2023. Generative Image AI Using Design Sketches as input: Opportunities and Challenges. In *Creativity and Cognition*. ACM, Virtual Event USA, 254–261. https://doi.org/10.1145/3591196.3596820

[132] Hongbo Zhang, Pei Chen, Xuelong Xie, Zhaoqu Jiang, Yifei Wu, Zejian Li, Xiaoyu Chen, and Lingyun Sun. 2025. FusionProtor: A Mixed-Prototype Tool for Component-level Physical-to-Virtual 3D Transition and Simulation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–19. https://doi.org/10.1145/3706598.3713686

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

[134] Hongbo Zhang, Pei Chen, Jingwen Yang, Yifei Wu, Zhaoqu Jiang, Xuelong Xie, Weitao You, and Lingyun Sun. 2025. IEDS: Exploring an Intelli-Embodied Design Space Combining Designer, AR, and GAI to Support Industrial Conceptual Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–25. https://doi.org/10.1145/3706598.3713528

[135] Lei Zhang, Jin Pan, Jacob Gettig, Steve Oney, and Anhong Guo. 2024. VRCopilot: Authoring 3D Layouts with Generative AI Models in VR. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, Pittsburgh PA USA, 1–13. https://doi.org/10.1145/3654777.3676451

A Appendix

This appendix provides methodological details, scope justification, decision rubrics, and tool-level scoring tables that support the main analyses.

A.1 Conceptual Contrast Sets for Scope Clarification

This section provides contrast sets to contextualize the scope of our review. These sets illustrate how (1) pre-2022 GenAI tools (2017–2021) and (2) 3D/AR/VR spatial interaction systems differ from the 2D, screen-based tools analyzed in this work. For clarity, we did not compute inter-rater agreement for these contrast sets, as they were conducted only within the final set of venues selected during our original venue-screening process.

A.1.1 Pre-2022 Contrast Set (2017–2021). Applying our original search query (with all exclusion terms unchanged; see Appendix A.2) to the period **2017–2021** returned **24** publications. A coarse title–abstract screen using the same inclusion criteria as our main review (implemented, user-facing 2D GenAI image-generation systems; see Section 3.1.3) yielded **one** qualifying system: *Creative Sketching Partner* [44]. This system provides model-guided transformations of user sketches to support ideation, but its output is limited to rough, stroke-based sketch variations rather than open-ended 2D image generation. Its interactions—sketch input with two sliders controlling visual and conceptual similarity—align partially with our frameworks (e.g., visual-based input, attribute-based input, global control, ideation, generators) but do not constitute a general-purpose image-generation workflow. Two additional systems surfaced [32, 42] but were excluded after full-text review: one focuses on poster-layout recomposition [32], and the other on fashion-collection matching and retrieval [42]. Neither performs generative 2D image creation. Under our existing search terms and eligibility criteria, extending the timeframe earlier than 2022 would therefore introduce only a very small number of systems, none of which align with the open-ended, user-facing 2D image-generation workflows examined in this review.

A.1.2 VR/AR/Spatial Contrast Set (2017–July 2025). Repeating our query **without VR/AR/3D exclusion terms** returned 332 papers (the original 286 plus **46** additional publications, including 24 from the 2017–2021 timeframe). A coarse screen surfaced **eight** potentially relevant papers; after removing three already covered in the pre-2022 contrast set and applying our inclusion criteria through full-text review, **two** systems qualified as implemented, user-facing generative spatial interfaces [115, 132]. Both systems primarily support **3D object or scene generation**, rather than 2D image generation. Including them would therefore require extending our inclusion criteria to encompass 3D asset-, prototype-, or scene-generation workflows. FusionProtor [132] is an MR interface centered on gesture-based selection, browsing, and manipulation of 3D objects, with menu interactions and text entry in mixed reality. DreamCrafter [115] is a VR-based 3D scene-editing system supporting direct object manipulation and voice-based global editing. These interactions introduce spatial dimensions—such as rotation, positioning, and scaling of 3D objects, as well as in-environment placement of generated content—that fall outside the scope of 2D, screen-based workflows. The remaining papers involved 3D modeling through standard 2D interfaces (e.g., chair modeling) [50], 3D sketching workflows executed on 2D screens [97], or conceptual or embodied design discussions without implemented systems [134], and were excluded accordingly. In summary, including AR/VR/MR modalities under our current search would surface only a minimal number of additional systems, and only if 3D generation were considered within scope beyond 2D image generation. The interaction paradigms observed—**embodied manipulation** of 3D objects, **viewpoint-dependent inspection**, and **in-environment spatial arrangement**—reflect the specific demands of 3D creation workflows and differ substantially from 2D, screen-based generative processes. Incorporating such systems would therefore require extending our framework to support these characteristics (e.g., spatial editing, gesture-based object control, and immersive feedback mechanisms).

A.2 Search Query Formulation Process

We followed ACM DL's advanced search guidelines to construct our search terms. This section shows our query development process for transparency and reproducibility.

A.2.1 Basic Search. We began with exact phrases and keywords such as "generative AI," "image generation," "interfaces," "HCI," and "design" to identify papers relevant to our research goals. The basic query used was:

```
("generative AI" OR "image generation") AND ("interfaces" OR
HCI OR design)
```

A.2.2 Expanded Search. We broadened the query by adding synonyms, acronyms, and related concepts (e.g., "AI-generated image," "generative model," "Generative Adversarial Network," and diffusion). To capture interface-related work, we included terms such as "tool," "creative support," "interaction design," and "usability."

A.2.3 Wildcard and Query Optimization. We used the wildcard (*) and OR operators to capture singular, plural, and related variations of key terms (e.g., interact*, creat*). Redundant terms were removed when covered by wildcards. This ensured broad coverage without unnecessary duplication.

A.2.4 Excluding Certain Topics. To filter out model-centric technical contributions, we excluded terms such as "code," "programming," and "software," as well as modality-specific or domain-specific terms outside our scope (e.g., "VR," "3D," "point cloud," "audio," "patient"). This refinement ensured that the retrieved results emphasized creativity, interface design, and user experience rather than technical or domain-specific advancements.

A.2.5 Finalized Queries. We applied two complementary queries to titles or abstracts, reflecting differences in how GenAI systems are described across papers.

Title Query (broader). This query was used to capture papers that explicitly label new systems or tools as GenAI-related or creative design systems.

```
("generative AI" OR "image generation" OR "AI-generated" OR
genAI OR "text-to-image" OR "image-to-image" OR "generative
model" OR "generative models" OR "Generative Adversarial Network"
OR diffusion OR GAN* OR "generative artificial intelligence"
OR "AI model" OR "AI models") AND ("user interface" OR interface*
OR "user experience" OR UX OR UI OR "user interaction" OR
interaction OR interact* OR "interaction design" OR usability
OR HCI OR "user study" OR "user studies" OR "user evaluation"
OR "user testing" OR "multi-modal" OR multimodal* OR tool*
OR "design tool" OR "creative support" OR "co-creation" OR
prototype OR design* OR creat* OR art* OR visual* OR express*
OR aesthetic*) AND NOT ("virtual reality" OR VR OR "augmented
reality" OR AR OR "mixed reality" OR MR OR "extended reality"
OR XR OR "3D" OR "point cloud" OR mesh OR "neural radiance"
OR music OR "sound design" OR audio OR vibration OR patient
OR medical OR health OR clinical OR code OR programming OR
software)
```

Abstract Query (narrower). This query was designed to capture studies describing interaction design, usability, and evaluation aspects, since abstracts often highlight study methods and results rather than tool names. Compared to the title query, we removed broad creative terms such as "visual" or "design" because in abstracts these words frequently appear in technical model papers (e.g., visual features, design of architectures) that are not aligned with our inclusion criteria. Keeping the abstract query narrower reduced false positives while still ensuring we retrieved relevant user-facing systems.

```
("generative AI" OR "image generation" OR "AI-generated" OR
genAI OR "text-to-image" OR "image-to-image" OR "generative
model" OR "generative models" OR "Generative Adversarial Network"
OR diffusion OR GAN* OR "generative artificial intelligence"
OR "AI model" OR "AI models") AND ("user interface" OR interface*
OR "user experience" OR UX OR UI OR "user interaction" OR
interaction OR interact* OR "interaction design" OR usability
OR HCI OR "user study" OR "user studies" OR "user evaluation"
OR "user testing") AND NOT ("virtual reality" OR VR OR "augmented
reality" OR AR OR "mixed reality" OR MR OR "extended reality"
OR XR OR "3D" OR "point cloud" OR mesh OR "neural radiance"
OR music OR "sound design" OR audio OR vibration OR patient
```

`OR medical OR health OR clinical OR code OR programming OR software)`

A.3 Decision Rubric for Attribute-Based Input vs. Parameter Control

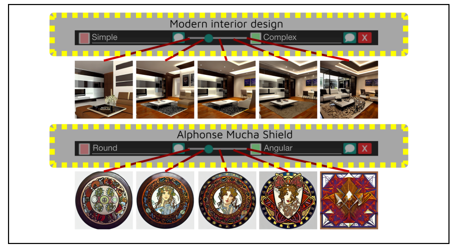
This appendix presents the decision rubric used to distinguish *attribute-based input* from *parameter control*. The distinction is grounded in **user-facing semantics**—that is, how controls are labeled and framed in the interface—rather than in implementation details. *Attribute-based inputs* adjust **perceptual, aesthetic, or semantic qualities** (e.g., color, lighting, style, composition) and are framed in design-oriented terms (e.g., "warmer lighting," "increase vibrancy," "more like this image"). These controls do not require an understanding of model parameters, even when they are implemented through model-level mechanisms. *Parameter controls* expose **settings at the model or inference level** (e.g., seed, sampling steps, denoising strength, model version) that directly influence the generative process. They are typically presented in numeric or system-level terms and often presume some familiarity with algorithmic behavior.

A.3.1 Rubric. Table 4 summarizes the operational criteria we used to classify controls and resolve borderline cases.

Table 4: Operational criteria distinguishing *Attribute-Based Input* and *Parameter Control*. The rubric clarifies how controls were classified and how borderline cases were resolved.

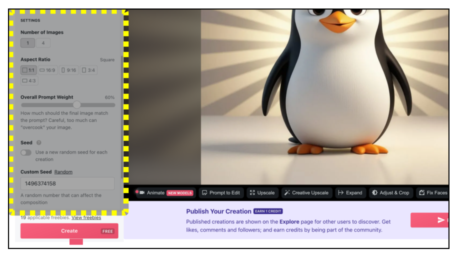
Criterion	Attribute-Based Input	Parameter Control
User-facing semantics	Adjust perceptual, aesthetic, or semantic qualities (e.g., style, lighting, composition).	Adjust system- or model-level variables (e.g., seed, sampling steps, denoising strength).
Labeling and framing	Design-oriented or descriptive terms (e.g., "warmer lighting," "increase vibrancy,") or similarity-based feedback ("more like this").	Numeric or system-level terms (e.g., "seed: 123," "steps: 50," "denoise: 0.7").
Interpretability	Requires no model-parameter knowledge; focuses on visual or semantic properties.	Typically requires familiarity with model configuration or inference behavior.
Functional purpose	Specifies *what* the output image should express (semantic or visual attributes).	Specifies *how* the model should operate (algorithmic or stochastic behavior).
Decision rule (borderline cases)	If framed as a semantic or aesthetic adjustment using design-oriented terms ⇒ classify as attribute-based.	If framed as model configuration or system-level tuning ⇒ classify as parameter control.

A.3.2 Examples of Attribute-Based Input and Parameter Control. Figure 7 provides representative examples illustrating the distinction between design-oriented attribute controls and system-level parameter controls.

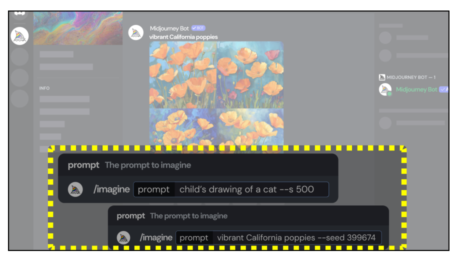


(a) PromptPaint [18]: adjusts visual attributes (e.g., style, form) through sliders framed as aesthetic controls rather than parameter settings.

(b) RoomDreaming [121]: supports perceptual adjustments (e.g., layout coherence) using design-direction sliders and "like" feedback, without exposing algorithmic parameters.



(c) NightCafe [70]: exposes model-level parameters (e.g., seed, prompt weight, model selection) that users adjust numerically to influence generation.



(d) Midjourney [67]: provides system-level controls (e.g., `--seed 399674`, `--s 500`) that configure generative variability and model behavior.

Figure 7: Examples of Attribute-Based Input (a–b) and Parameter Control (c–d). Panels (a) and (b) show design-oriented semantic or perceptual adjustments, while panels (c) and (d) show numeric or system-level settings that configure model- or inference-level behavior. Panels (a) and (b) are adapted from ACM-published works (© ACM). Panels (c) and (d) show screenshots of commercial GenAI interfaces (© NightCafe; © Midjourney).

A.4 Tool Support within Frameworks

A.4.1 Support for Interaction Methods. Table 5 reports support levels across input types, control levels, and interaction styles for all tools in our dataset.

Table 5: GenAI tools categorized by interaction methods, where ●●● denotes strong support, ●● denotes moderate support, ● denotes limited support, and – denotes no support.

	Tool Name	Input Types				Control Levels			Interaction Styles			
		Text-Based	Visual-Based	Attribute-Based	Parameter Control	Global Control	Component Control	Precision Control	Direct Manipulation	GUI Interactions	Parallel Exploration	Iterative Adjustments
Academic Tools	AIdeation [122]	●●●	●	●●	–	●●●	●●	–	–	●●	●●●	●●
	AutoSpark [14]	●●●	●●	●●	–	●●●	●●	–	–	●●●	●●●	●●
	Brickify [99]	●●	●●●	●●	–	●●●	●●●	●	●●●	●●	●●	●●●
	CreativeBlends [107]	●●●	●	●●	–	●●	●●	–	–	●●●	●●●	●●
	CreativeConnect [17]	●●	●●●	●●	–	●●●	●	–	●	●●	●●●	●
	DesignPrompt [82]	●●●	●●	●●	–	●●●	●●●	●	●	●●	●●●	●●
	DesignWeaver [111]	●●●	●	●●●	–	●●●	●	–	–	●●	●●●	●●
	Drawing-in-Steps [125]	●●●	●●●	●	●	●●●	●●●	●	●●	●●	●	●●
	DreamSheet [3]	●●●	–	●	●	●●●	●	–	–	●●	●●●	●●
	FusAIn [83]	●●●	●●●	●●	●	●●●	●●●	●●●	●●	●●	●●	●●●
	GanCollage [119]	●●	●●	●●	●	●	●	–	●	●	●●●	●
	generative.fashion [23]	–	●	●●●	●	●	●	–	●	●●	●●●	●
	GenQuery [103]	●●●	●●●	●	–	●●●	●●	●	●	●●	●●●	●●
	Inkspire [55]	●●	●●●	●●	●	●●●	●●	●	●●	●●	●●	●●
	Opal [59]	●●●	–	●●	–	●●●	–	–	–	●●	●●●	●
	Paratrouper [51]	●●●	●●●	●	●●	●●●	●●●	●●	●●	●●	●●●	●●●
	PlantoGraphy [38]	●●●	●	●●●	–	●●●	●●	–	–	●●●	●●	●
	PromptCharm [124]	●●●	●●	●●	●	●●●	●●●	●	●	●●●	●●	●●●
	Promptify [9]	●●●	–	●	●	●●●	–	–	●	●●●	●●●	●
	PromptMap [1]	●●●	–	●●	–	●●●	–	–	–	●●●	●●	●
	PromptPaint [18]	●●●	●●●	●●●	●	●●●	●●●	●●	●●	●●●	●●	●●●
	ProtoDreamer [133]	●●●	●●	●	●	●●	●●●	●●	●●	●●	●●●	●●●
	RoomDreaming [121]	●	●	●●	●	●●	●	–	–	●●	●●●	●
	SketchFlex [56]	●●	●●●	●●	●	●●●	●●●	●	●●	●●	●●	●●
	StyleMe [127]	–	●●●	–	–	●	●●	●	●	●	●	●●
	StyleWe [126]	–	●●●	–	–	●	●●	●	●	●	●	●●
	Varif.ai [66]	●●●	–	●●●	●	●●●	–	–	–	●●●	●●●	●●
	WorldSmith [22]	●●●	●●	–	–	●●●	●●●	●	●●	●●	●●●	●●●
Commercial Tools	DALL-E 3 [74]	●●●	●	●	●	●●●	●●	●	●	●	●	●●
	DreamStudio [25]	●●●	●●	●	●●	●●●	●●	●	●	●●	●●	●●
	Firefly [2]	●●●	●●●	●●●	●	●●●	●●●	●●●	●●	●●●	●●	●●●
	Midjourney [67]	●●●	●●	●●●	●●●	●●●	●●	●	●	●	●●	●●
	NightCafe [70]	●●●	●●	●●	●●	●●●	●●	●	●	●●	●●	●●
	RunwayML [94]	●●●	●●	●	●●	●●●	●●	●	●	●●	●●	●●
	StarryAI [105]	●●●	●●	●●	●●	●●●	●●	●	●	●●	●●	●●
	VisualElectric [117]	●●●	●●	●●	●	●●●	●●	●	●	●●	●●●	●●
	Vizcom [118]	●●●	●●●	●●	●	●●	●●●	●●●	●●	●●●	●●●	●●●

A.4.2 Support for Creative Processes. Table 6 summarizes tool support across the four creative process phases: ideation, exploration, refinement, and evaluation.

Table 6: GenAI tools categorized by creative processes, where ●●● denotes strong support, ●● denotes moderate support, ● denotes limited support, and – denotes no support.

	Tool Name	Ideation	Exploration	Refinement	Evaluation
Academic Tools	AIdeation [122]	●●●	●●●	●	●
	AutoSpark [14]	●●●	●●●	●●	●●●
	Brickify [99]	●●●	●●	●●●	●
	CreativeBlends [107]	●●●	●●●	●	●●
	CreativeConnect [17]	●●●	●●	●	●
	DesignPrompt [82]	●●●	●●●	●●	●
	DesignWeaver [111]	●●●	●●●	●●	●
	Drawing-in-Steps [125]	●●	●●●	●●	●
	DreamSheet [3]	●●●	●●●	●	●●
	FusAIn [83]	●●●	●●	●●●	●
	GanCollage [119]	●●●	●●●	●	●
	generative.fashion [23]	●●	●●●	●	●●
	GenQuery [103]	●●●	●●●	●●	●●
	Inkspire [55]	●●●	●●	●●	●
	Opal [59]	●●●	●●	●	●
	Paratrouper [51]	●●●	●●●	●●	●●
	PlantoGraphy [38]	●●●	●●●	●	●
	PromptCharm [124]	●●●	●●●	●●	●●
	Promptify [9]	●●●	●●	●	●●
	PromptMap [1]	●●●	●●	●	●●
	PromptPaint [18]	●●●	●●●	●●●	●
	ProtoDreamer [133]	●●	●●●	●●●	●●●
	RoomDreaming [121]	●●●	●●●	●	●●
	SketchFlex [56]	●●●	●●	●●●	●
	StyleMe [127]	●●	●●●	●●	●
	StyleWe [126]	●●	●●●	●●	●
	Varif.ai [66]	●●●	●●●	●●	●●●
	WorldSmith [22]	●●	●●	●●	●●
Commercial Tools	DALL·E 3 [74]	●●●	●●●	●●	●
	DreamStudio [25]	●●●	●●●	●●	●
	Firefly [2]	●●●	●●●	●●●	●●
	Midjourney [67]	●●●	●●●	●●	●
	NightCafe [70]	●●●	●●●	●●	●
	RunwayML [94]	●●●	●●●	●●	●
	StarryAI [105]	●●●	●●●	●●	●
	VisualElectric [117]	●●●	●●●	●●	●
	Vizcom [118]	●●●	●●●	●●●	●●

A.4.3 Support for Tool Functionalities. Table 7 reports support levels across the five functionality categories used in our analysis.

Table 7: GenAI tools categorized by tool functionalities, where ●●● denotes strong support, ●● denotes moderate support, ● denotes limited support, and – denotes no support.

	Tool Name	Generators	Editors	Blenders	Variation Explorers	Prompt Guides
Academic Tools	AIdeation [122]	●●●	●	●	●●●	●●●
	AutoSpark [14]	●●●	●	–	●●●	●●●
	Brickify [99]	●●●	●●●	●●	●●	●●
	CreativeBlends [107]	●●●	●	●●●	●●●	●●●
	CreativeConnect [17]	●●●	–	–	●●	●●●
	DesignPrompt [82]	●●●	●●	●	●●●	●●●
	DesignWeaver [111]	●●●	●	●	●●●	●●●
	Drawing-in-Steps [125]	●●●	●●	●	●●	●
	DreamSheet [3]	●●●	●	●	●●●	●
	FusAIn [83]	●●●	●●●	●●	●●	●
	GanCollage [119]	●●●	●	●●	●●●	●●
	generative.fashion [23]	●●	●	●●	●●●	●
	GenQuery [103]	●●●	●●	●	●●●	●●●
	Inkspire [55]	●●●	●●	–	●●	●●
	Opal [59]	●●●	–	–	●●●	●●●
	Paratrouper [51]	●●●	●●●	●●●	●●●	●
	PlantoGraphy [38]	●●●	●	–	●●●	●
	PromptCharm [124]	●●●	●●	–	●●●	●●●
	Promptify [9]	●●●	–	–	●●●	●●●
	PromptMap [1]	●●●	–	–	●●●	●●●
	PromptPaint [18]	●●●	●●●	●●	●●●	●
	ProtoDreamer [133]	●●●	●●●	●	●●●	–
	RoomDreaming [121]	●●●	–	–	●●●	–
	SketchFlex [56]	●●●	●●	–	●●	●●●
	StyleMe [127]	●●	●●	●	●●	–
	StyleWe [126]	●●●	●●	●●	●●	–
	Varif.ai [66]	●●●	–	●●	●●●	●●
	WorldSmith [22]	●●●	●●	●	●●●	–
Commercial Tools	DALL·E 3 [74]	●●●	●●	–	●●	–
	DreamStudio [25]	●●●	●●	–	●●●	●
	Firefly [2]	●●●	●●●	●●●	●●●	●
	Midjourney [67]	●●●	●●	●●●	●●●	●
	NightCafe [70]	●●●	●●	–	●●●	–
	RunwayML [94]	●●●	●●	●	●●●	●
	StarryAI [105]	●●●	●●	–	●●●	–
	VisualElectric [117]	●●●	●●	●	●●●	●●
	Vizcom [118]	●●●	●●●	●●●	●●●	●

A.5 Tool Examples by Support Level

This appendix provides representative tools illustrating strong (3), moderate (2), and limited (1) support for each framework dimension. Tools with a score of 0 are omitted because they indicate the absence of features. All scoring details and evaluation rationales are provided in the supplementary material.

A.5.1 Interaction Methods. Table 8 presents triad examples (scores 3–1) for each interaction-method dimension, illustrating how our scoring criteria distinguish levels of support.

Table 8: Triad examples for the Interaction Methods framework.

Interaction Methods	Strong Support (●●●)	Moderate Support (●●)	Limited Support (●)
Input Types			
Text-Based Input	**Promptify** [9]: Supports flexible natural-language prompts (any length, negatives) with editable suggestions.	**Brickify** [99]: Links text to visual tokens via "concept markers," enabling localized effects but restricting expressiveness due to menu-driven entry and the lack of a free-form prompt box.	**RoomDreaming** [121]: Accepts optional text keywords and preference signals ("like") that nudge stylistic direction but do not drive full prompt-based generation.
Visual-Based Input	**Vizcom** [118]: Enables sketching and drawing directly on the canvas with adjustable influence and layered editing, supporting precise visual intent expression.	**DesignPrompt** [82]: Allows image uploads and reference-based color selection but limits drawing to simple inpainting, reducing support for detailed visual editing.	**AIdeation** [122]: Includes visual search and uploads for ideation, yet provides no tools for direct visual editing or sketch-driven generation.
Attribute-Based Input	**PromptPaint** [18]: Offers fine-grained sliders for blending and weighting stylistic attributes, enabling nuanced control over semantic and visual properties.	**CreativeConnect** [17]: Allows selection of themes, moods, or abstract concepts but lacks tuning or weighting options for precise adjustments.	**DALL·E 3** [74]: Provides fixed presets for aspects such as style or aspect ratio, offering only coarse stylistic variability.
Parameter Control	**Midjourney** [67]: Exposes a wide range of model-level parameters (e.g., seed, chaos, model version) that allow granular configuration of generation behavior.	**NightCafe** [70]: Supports several basic numeric settings (resolution, aspect ratio, iteration steps) but omits deeper inference-level controls.	**Drawing-in-Steps** [125]: Provides only a single refinement-intensity slider, offering minimal parameter-control functionality.
Control Levels			
Global Control	**CreativeConnect** [17]: Recombines extracted keywords to modify themes, poses, and conceptual intent, enabling broad high-level steering.	**ProtoDreamer** [133]: Generates variants from a prototype combined with prompts and presets, but global variation remains constrained by the original form.	**generative.fashion** [23]: Relies on predefined modes and basic attribute sliders, limiting global control to preset combinations.
Component Control	**Brickify** [99]: Supports direct manipulation of decomposed tokens (resize, move, recolor, link), enabling explicit component-level editing.	**AIdeation** [122]: Adjusts object attributes through text or reference images, but control is indirect and lacks precise spatial manipulation.	**generative.fashion** [23]: Adjusts clothing components via latent-space axes (e.g., sleeve length, color) but cannot introduce new components or modify layout structure.
Precision Control	**Vizcom** [118]: Provides strong regional precision via drawing, selection, erasing, and layered tools for controlled fine-grained refinement.	**Paratrouper** [51]: Uses masks and brushes for local edits, though with fewer precision tools and less layering flexibility than high-precision systems.	**Brickify** [99]: Supports basic spatial refinement through token resizing/grouping but offers no mask or brush tools for detailed adjustments.
Interaction Styles			
Direct Manipulation	**Brickify** [99]: Offers direct visual manipulation of tokens (drag, resize, recolor, group) to express intent without text.	**ProtoDreamer** [133]: Enables targeted object adjustments using anchors and arrows but lacks drawing or free-form manipulation.	**Promptify** [9]: Provides drag/zoom interactions for moodboard layout only, not for image editing.
GUI Interactions	**PlantoGraphy** [38]: Includes scene graphs, layout panels, and attribute sliders that support detailed GUI-driven scene specification (e.g., weather, time of day).	**GenQuery** [103]: Provides a multimodal GUI with search and clickable elements, but lacks advanced widgets for deeper control.	**Midjourney** (Discord) [67]: Command-driven text interface with minimal GUI elements; limited visual interactivity.
Parallel Exploration	**AIdeation** [122]: Offers multiple variations, 8-image comparison views, history, and regeneration for broad exploration.	**SketchFlex** [56]: Generates several object-level alternatives but does not support full-scene or grid-based branching.	**DALL·E 3** [74]: Presents output history in a scrollable list without structured side-by-side comparison tools.
Iterative Adjustments	**Paratrouper** [51]: Supports iterative refinement via prompt edits, seed variation, local adjustments, and history, with localized controls supporting strong image-level iteration.	**Varif.ai** [66]: Supports iterative cycles (generate, check diversity, adjust, regenerate) with revert/branch history; adjustments operate at the attribute or distribution level, with no local image or region edits.	**Promptify** [9]: Allows prompt-level iteration only (e.g., modifying keywords, tones, or styles through reprompting), with no per-image or region-level adjustments.

A.5.2 Creative Processes. Table 9 presents triad examples (scores 3–1) for each creative process phase.

Table 9: Triad examples for the Creative Processes framework.

Creative Processes	Strong Support (●●●)	Moderate Support (●●)	Limited Support (●)
Ideation	**CreativeConnect** [17]: Extracts keywords from images and recombines them to generate new concepts, supporting broad, semantically guided ideation.	**ProtoDreamer** [133]: Produces idea variations from an uploaded prototype with prompts and style presets, but the generated space remains closely tied to the input form.	No tool in our corpus demonstrated limited ideation support.
Exploration	**GenQuery** [103]: Enables wide visual exploration through search, replace, and modify operations on selected images or regions, combined with text or reference-image adjustments.	**Opal** [59]: Offers thematic exploration via text and keyword recommendations but does not support branching or multi-variant exploration from a single image.	No tool in our corpus demonstrated limited exploration support.
Refinement	**FusAIn** [83]: Provides iterative pen-based refinement, allowing users to reapply strokes and decompose or recombine elements, making refinement a core system capability.	**WorldSmith** [22]: Supports regional adjustments via inpainting (region + text) with tile-level editing through a tree view, though refinement remains primarily inpainting-focused.	**Promptify** [9]: Offers keyword and tone manipulations for prompt-level refinement but lacks region selection or localized editing.
Evaluation	**AutoSpark** [14]: Supports detailed evaluation through side-by-side image and text comparisons, heatmap overlays, and parameter visualizations that enable fine-grained assessment.	**PromptCharm** [124]: Provides version control and comparison views but lacks structured evaluation metrics, relying mainly on manual side-by-side judgment.	**Midjourney** [67]: Displays multiple generations in a batch, yet evaluation remains manual and unstructured with no comparative interfaces or metrics.

A.5.3 Tool Functionalities. Table 10 presents triad examples (scores 3–1) for each functionality category.

Table 10: Triad examples for the Tool Functionalities framework.

Tool Functionalities	Strong Support (●●●)	Moderate Support (●●)	Limited Support (●)
Generators	**Midjourney** [67]: Produces high-resolution images with consistent stylistic coherence and supports a wide range of prompt-driven generation behaviors.	**generative.fashion** [23]: Generates designs through random sampling and latent blending, but allows user steering only after initial outputs.	No tool in our corpus demonstrated limited generator functionality.
Editors	**Vizcom** [118]: Supports detailed local refinement through drawing, selection, and layer-based editing, enabling precise region-level adjustments.	**DALL·E 3** [74]: Allows region-based inpainting by selecting an area and providing text, but lacks drawing tools or multi-layer editing.	**DreamSheet** [3]: Performs "edits" only through spreadsheet parameters (e.g., seed, CFG) or text prompts, without region-level capabilities.
Blenders	**CreativeBlends** [107]: Supports multi-object fusion using shared attributes (e.g., combining semantic concepts) as a primary interaction mechanism.	**Brickify** [99]: Blends at the component level by decomposing reference images into tokens that users recombine, but cannot blend entire images.	**Drawing-in-Steps** [125]: Sequentially combines user sketches and text prompts but does not blend finished images or external references.
Variation Explorers	**AutoSpark** [14]: Offers comprehensive exploration using node-based visualizations and branching paths that support exploratory ideation.	**SketchFlex** [56]: Generates and replaces object-level candidates iteratively, though exploration remains linear and object-focused.	No tool in our corpus demonstrated limited support for variation exploration.
Prompt Guides	**PromptCharm** [124]: Maps prompt words to image regions and suggests alternatives, enabling guided prompt construction with minimal effort.	**VisualElectric** [117]: Provides predictive suggestions and preset style options, offering helpful but limited prompt guidance.	**PlantoGraphy** [38]: Supplies predefined style keywords that help users understand domain-specific styles but do not support interactive prompt refinement.