ImaginationVellum: Generative-AI Ideation Canvas with Spatial Prompts, Generative Strokes, and Ideation History

Nicolai Marquardt
Microsoft Research
Redmond, WA, USA
nicmarquardt@microsoft.com

Asta Roseway
Microsoft Research
Redmond, WA, USA
houseofroseway@gmail.com

Hugo Romat
Microsoft Research
Redmond, WA, USA
romathugo@microsoft.com

Payod Panda
Microsoft Research
Cambridge, United Kingdom
panda@payodpanda.com

Michel Pahud
Microsoft Research
Redmond, WA, USA
mpahud@microsoft.com

Gonzalo Ramos
Microsoft Research
Redmond, WA, USA
gonzo.ramos@gmail.com

Steven M. Drucker
Microsoft Research
Redmond, WA, USA
sdrucker@gmail.com

Andrew D. Wilson
Microsoft Research
Redmond, WA, USA
awilson@microsoft.com

Ken Hinckley
Microsoft Research
Redmond, WA, USA
kenneth.p.hinckley@gmail.com

Nathalie Riche
Microsoft Research
Redmond, WA, USA
nath@microsoft.com

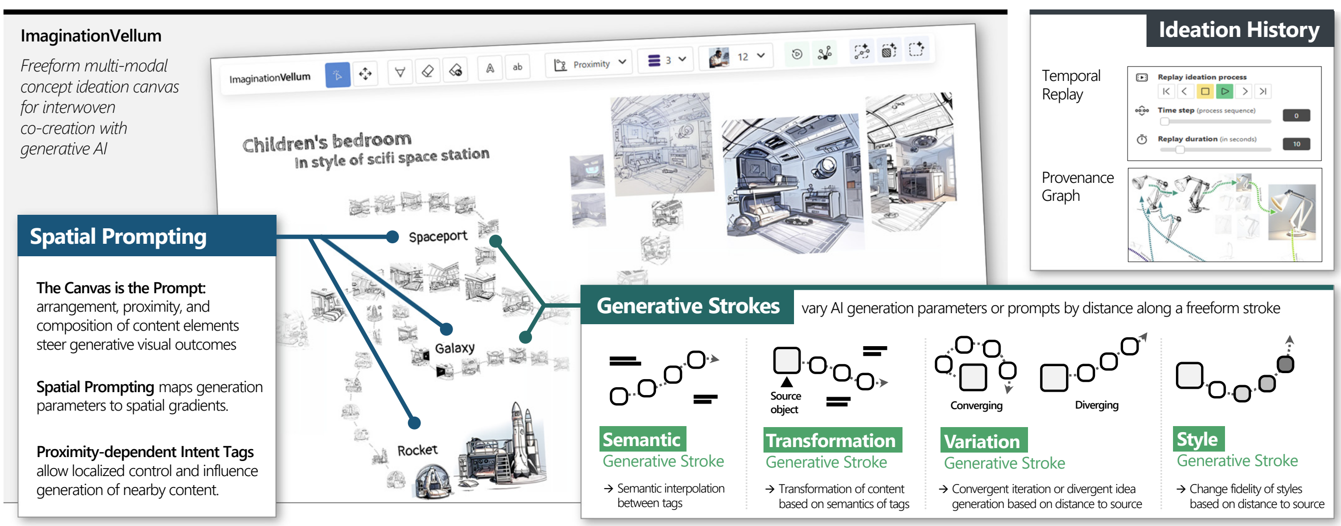

Figure 1: ImaginationVellum leverages spatial prompting for iterative visual co-creation with generative AI.

Abstract

We introduce ImaginationVellum, a multi-modal spatial canvas for early-stage visual ideation and concept sketching with generative AI. The resulting system supports a unique style of human-AI co-creation where *the canvas is the prompt*. This means that ImaginationVellum employs the entire 2D canvas as an active prompt space, where spatial arrangement, proximity, and composition of diverse content elements—inking, text, images, and intermediate results—steer generative visual outcomes. As a technical probe, ImaginationVellum contributes a set of spatially-grounded direct manipulation tools for iterative visual ideation. In particular, we introduce Generative Strokes—freeform strokes that spatially modulate generation and prompt-parameters (articulated along multiple latent semantic or stylistic dimensions). These techniques afford rapid traversal of design spaces via convergence, divergence, re-composition, blending, and remixing of concepts. We detail the system architecture, design rationale, proximity-dependent intent tags for localized control, and methods for spatial prompting and varying output along spatial gradients. Temporal replay and visualization of provenance make ideation trajectories actionable, turning the design process itself into an artifact that supports reflection-in-action and revisitation of design decisions. We report insights

UIST '25, Busan, Republic of Korea

ACM ISBN 979-8-4007-2037-6/25/09
https://doi.org/10.1145/3746059.3747631

from a preliminary study of how users construct, steer, and revisit ideas using spatial prompts, and discuss tradeoffs in modulating spatially-dependent content generation.

CCS Concepts

• Human-centered computing → Interactive systems and tools; **Interaction paradigms**; **• Computing methodologies → Artificial intelligence**.

Keywords

spatial prompting, generative strokes, generative AI, human-AI co-creation, sketching, spatial ideation canvas

ACM Reference Format:
Nicolai Marquardt, Asta Roseway, Hugo Romat, Payod Panda, Michel Pahud, Gonzalo Ramos, Steven M. Drucker, Andrew D. Wilson, Ken Hinckley, and Nathalie Riche. 2025. ImaginationVellum: Generative-AI Ideation Canvas with Spatial Prompts, Generative Strokes, and Ideation History. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25), September 28–October 01, 2025, Busan, Republic of Korea.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3746059.3747631

1 Introduction

ImaginationVellum is a freeform concept ideation canvas that supports early-stage design sketching via co-creation with generative AI, as illustrated in Figure 1, such that ***the canvas is the prompt***. The ability to freely arrange and organize content—as well as intermediate work-artifacts—is an important way in which people make "intelligent use of space" [46] to externalize thinking and structure complex activities in the spatial arrangement of their work [38, 99]. ImaginationVellum makes this spatial intelligence bidirectional by supporting natural human creative practices and cognitive processes while also enabling spatial arrangement, proximity, and sketching around and on top of content as expressive inputs to generative AI for human-AI co-creation.

Linear, text-centric chat, however, lacks such spatial affordances as an interface to generative AI models, among other challenges that users face with prompting [13, 95, 109]. ImaginationVellum thus serves as a technology probe [39] that makes key technical and interaction-technique contributions towards a user experience for design sketching that leverages this key concept that *the canvas is the prompt.*

Sketching as a design practice supports creative workflows [15, 17, 26, 72] by externalizing early-stage concepts, making it an important bridge from the spark of an idea to the broader exploration of its potential [23, 27, 83, 84]. To support design sketching, ImaginationVellum's pan-and-zoom 2D canvas supports mixed (multi-modal) object types including ink strokes, text, images, and generated content, all of which—including the composition or select portions thereof—can be freely arranged, layered, juxtaposed, or interpolated between as "inputs" to generate further content. Through ***Spatial Prompting***, the ImaginationVellum canvas modulates generative behavior based on spatial relationships between items, and adapts prompts and generation parameters based on spatial gradients. Expanding on earlier notions of spatial prompting [24], this allows for iterative construction, decomposition, and reinterpretation of prompts directly within the creative ideation environment of the canvas.

While ImaginationVellum leverages freeform sketching via pen + touch, if available on a device, it also supports mouse and keyboard input—and does not necessarily require manual sketching expertise. Indeed, ImaginationVellum preferentially generates its co-created outputs as sketch-style images: the rapid, low-fidelity character of sketches leaves room for the imagination [8, 17] and invites reflection-in-action (that is, a "conversation" between the designer and the creative re-interpretation of their materials) [87]. ImaginationVellum thus contributes a 2D freeform canvas that bolsters these practices by using both the content and the spatial properties of the canvas itself to steer human-AI co-creation.

To support the ideation process and empower AI to make intelligent use of space, ImaginationVellum introduces a set of spatially grounded tools that steer the generation of new visual content through spatial prompting. In particular, we introduce ***Generative Strokes*** as a class of tools that varies AI generation parameters by distance along a freeform stroke. Users can create strokes along multiple latent dimensions, such as semantic interpolation between intent tags, transformation of selected content, convergent iteration and refinement, divergent idea generation, or between low- and high-fidelity visual styles. Feedforward techniques graphically indicate the impact of nearby content elements on the output of Generative Strokes, so that users can see how and to what extent nearby content influences a generated output. Generative Strokes leverage *Proximity-dependent Intent Tags* that weigh elements incorporated into the automatically constructed prompt according to how close they are to the generated content. These can be text snippets that users create to identify a concept (e.g. "Sci-Fi Lamp" or "blocky, sharp, spikes") but proximity also influences other model parameters (e.g., temperature, diffusion model weights) according to spatial distance. The canvas furthermore preserves the ***history of the design process as an artifact***, making the user's prompt history, visual alternatives, and trajectories of ideation available for review and re-activation. First, *Temporal Replay* of the design-sketch timeline lets the designer rewind their ideation steps and choose any moment as a fresh point of departure. And second, *Provenance Graphs* allow the designer to quickly see how all content elements on the canvas contributed to the generation of a selected item.

These tools enable fluid transitions between *divergent* and *convergent* phases of design-sketching exploration—with support for branching styles and pursuit of multiple design possibilities in parallel—but also iteration and refinement of promising ideas by re-composing or re-combining select canvas elements. By their integration in ImaginationVellum's technical architecture, this paper makes the following contributions:

- Technical design choices supporting spatial prompting for generative AI, including mapping of model generation parameters to spatial gradients, to realize a human-AI co-creation experience where *the canvas is the prompt.*
- The ImaginationVellum system as a technology probe for exploring tightly-coupled human-AI visual co-ideation with spatially-aware canvas interaction techniques. In particular, *Generative Strokes* that allow modulating generative outcomes

across multiple latent semantic, convergent/divergent, and stylistic dimensions—for rapid exploration of design spaces, remixing, and blending of concept ideas directly on the canvas. We also introduce *Temporal Replay* and *Provenance Visualization* techniques that allow users to preserve their design process as an artifact.
- Preliminary user-study surfacing early insights about a higher sense of control when steering results via Generative Strokes and spatial prompting; as well as reactions to revisiting and revealing the human-AI ideation process, and differences in perceived values.

Taken together, these contributions demonstrate the rich and largely untapped design opportunities of a spatial canvas imbued with content and tools that make intelligent, proximity-dependent use of space for closely-coupled co-creation with generative AI. Drawing on the complementary strengths and capabilities of humans and AI, we believe experiences that appropriately design for and interweave these capacities could greatly accelerate and extend human design practices in exciting new ways.

In the following sections we first discuss key related work about ideation with generative AI, expressive tools to steer content generation, and use of spatial canvas ideation spaces. We then illustrate the expressive tools of ImaginationVellum with two walk-through scenarios. Next, we explain the design principles and core functionality of ImaginationVellum, and focus on the spatial prompts, Generative Strokes (across multiple dimensions of expression), the preservation of the history of the ideation process, and describe our technical implementation. We then present the findings of our user study with insights about the use of the spatial canvas interaction techniques and revisiting the process of ideation. We close with a discussion and suggest directions for future work.

2 Related Work

Our work builds on concepts in the literature at the intersection of design ideation, creativity support tools (CSTs), and the emergence of generative AI as a tool, design material, and medium.

2.1 Creativity, Ideation, & Freeform Sketching

Creativity and ideation are fundamentally non-linear processes—often characterized in design as a double-diamond structure of divergence, reflection, and convergence [17, 22]—that encompass a range of human activity [102] including embodied and tool-mediated actions externalized in the workspace itself [36–38, 46].

Creative activity has long favored freeform expression on 2D spatial work-surfaces (e.g. [16, 17, 26]). Sketchpad [97] pioneered direct manipulation with digital sketching, "eliminating typed statements"—a prescient glimpse of our desire in ImaginationVellum to move away from text-chat for prompting AI image generation. Freeform expression naturally fosters externalization of nascent ideas [29, 41, 83], enabling low-fidelity, ambiguous exploration [17, 33]. Previous work shows how digital inking can facilitate sketching [43, 93], support active reading and notetaking [35, 86], and afford other rich interactive capabilities [50, 85, 89].

In particular, canvas-based interfaces offer flexible spatial workspaces for organizing, composing, and reasoning with content. Systems such as Tivoli [68, 76], DigitalDesk [106], DynaPad [10], and others [19, 70] demonstrate how spatial arrangement supports creative knowledge work and contextual linking of media [79, 86]. Canvases can be "infinite" [11], but in practice many systems favor page-based [69, 108, 110] or otherwise constrained navigation [2, 65, 71]. Most importantly, ImaginationVellum demonstrates how generative AI can amplify early-stage design ideation on a digital canvas, particularly when the canvas's content (and the spatial arrangement thereof) serves as the prompt for human-AI co-creation.

Creativity Support Tools (CSTs) can provide technological support across the entire life-cycle of ideation, including asset creation, divergence, convergence, iteration, and managing the creative process itself [75, 90, 91]. Support for divergent thinking [17, 34, 98] is critical, but as ideas percolate, the designer must *reflect-in-action* on intermediate work-artifacts to converge on the next design move(s) to make [88], exemplified by systems such as SketchStorm [55] and IdeaHound [92]. ImaginationVellum offers a new form of CST for an AI-infused, freeform workspace where *the canvas is the prompt*.

2.2 Computational and AI-Assisted Sketching

Another form of artifact support looks to bring sketching skills to a wider audience, so that anyone can sketch creatively. Previous strategies include, for example, the close integration of sketch tutorial systems directly within apps [28], automatically inferring rich 3D objects from 2D scribbles [40], or interactive dynamic assistance lines and grids for 3D sketching [7]. Sketching support can take the form of a collaborative experience with a machine counterpart. Systems like DuetDraw [73] and Reframer [51] explored minimal additions and alterations done to a person's sketch by the AI in real time. And Inkspire [54] leverages scaffolding of a person's loose sketch strokes for product design exploration.

Traditionally, digital canvases as a medium (and tools to work on them) emphasize hand-drawn artifacts are draft- or sketch-like in nature. With the emergence of diffusion models [3], translating the description of ideas into renderings (at arbitrary levels of style or fidelity) has become a basic capability in the hands of many. With the emergence of generative AI, opportunities and challenges arise for a new generation of CSTs. Our work advances these concepts by building on the durable and battle-tested 2D canvas form factor.

Many generative AI tools generate images from a text prompt [63, 94], with changes enacted by reformulating the prompt. Emerging tools like Apple's Image Wand can generate images inside a 2D note-taking canvas, while Adobe's Project Concept extends Design Galleries [59] to support iterative image-drafting processes with *"richer ways to express and verify intent"* [75]. These advancements show how AI can mediate between sketches and prompts as co-existing design materials [49], hinting at inherently multi-modal experiences [24, 57]. ImaginationVellum explores not only the integration of generative capabilities into a design-ideation canvas, but also how the medium invites richer, more expressive interactions that support and align with existing creative processes.

2.3 Interacting with Generative AI beyond Traditional Prompts

Prompt-based interaction remains the dominant paradigm for steering generative models, but its ubiquity comes with challenges. Users

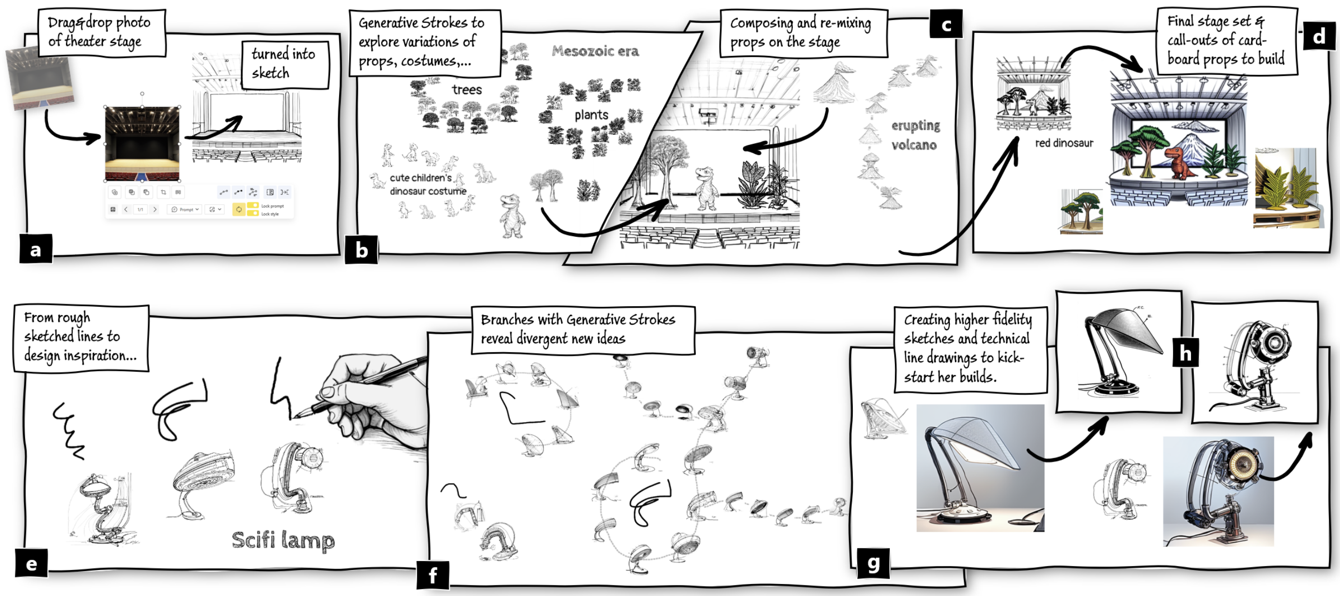

Figure 2: Walkthrough scenarios: (a–d) a teacher planning the stage setup for an elementary school theater production, and (e–h) a maker and builder envisioning a custom designed metal lamp inspired by science fiction.

often rely on trial-and-error strategies, struggle with intent formulation [13, 95, 109], and have a misaligned understanding of the capabilities of a model [58]. As ideation canvases incorporate generative AI, opportunities arise for user experiences that go beyond linear-text chat in terms of expressiveness and specificity. For example, recent systems augment prompting with scaffolding [52], visual feedback [104], and direct manipulation techniques [60].

Techniques also couple steering with tools to manipulate generation trajectories across temporal and semantic spaces, as in TaleBrush [21] to craft expressive narrative story arcs through brush strokes. And other expression techniques map out prompt possibilities. For example, PromptAid [66] uses visual analytics techniques to compare, test, and refine prompts. Even more granular, PromptPaint [20] and Prompt Fragments in AI-Instruments [82] expose the latent structure of prompt spaces, enabling users to explore alternative phrasings or generation paths. And FusAIn [78] affords fine-grained control of designers creations through loading expressive textures or colors into pens reconstructing attributes for visual prompts.

2.3.1 Spatial Layout and Canvases for AI. Space carries meaning on a canvas, and indeed spatial layout has been explored as a modality for steering generation. DesignPrompt [77] lets users spatially compose visual prompts by arranging images, text, and colors, offering an example of how layout can ground intent while offering intuitive and interpretable control over generative outputs. WorldSmith [24] and Transformative Lenses [82] allow users to guide content synthesis through region-based sketching or layout composition, modulating the diffusion process through local visual constraints. Similarly, CanvasPic [105] changes the attributes of generated images based on changes in layout, while users can place Intent Tags [32] in categorical regions to guide content generation.

While used to express intent, canvases remain a reflective *output* space, as seen in Sensecape [96], which uses a pan-and-zoom canvas to facilitate information foraging and synthesis with language models. Graphologue [42] offers a framework based on cells, generators, and lenses [45] that use the canvas to externalize the behavior of the model and allow for structured exploration of prompts.

By contrast, with ImaginationVellum, we use the canvas as an *input* surface to steer AI generation in nuanced ways. Inspired by visual ideation and sketching tools like SketchStorm [55] and Inkeraction [89], we aim for fluidity between generating, arranging, re-mixing, and re-prompting content. Our approach blurs the boundary between control surface [5, 6, 20, 21, 107] and visual ideation canvas [55, 89], promoting spatial composition as both a generative input and evolving artifact for output.

2.3.2 Exploring and Navigating Latent Space. To illuminate the near-infinite possibilities of generative AI [80], other strategies support exploring and navigating the latent space defined by particular models. Early systems such as Design Galleries [59] allowed users to explore (parametric) design variations via spatial layouts around an image, but more recent work rekindles notions of AI latent space navigation. For example, DreamLens [61] visualizes clusters of alternatives across design parameters, while ImageSense [47] and SemanticCollage [48] facilitate browsing and refinement of visual ideation through a semantically clustered, moodboard-like layout.

2.3.3 Visual representations of generative AI outcomes. To facilitate prompt discovery and refinement, systems such as Promptify [14] cluster and display generated images in chosen dimensions, and incorporate prompt suggestions. DreamSheets [4] allows the rapid sampling of image prompt combinations in a spreadsheet-like interface. Further possibilities for exposing latent spaces include interactive visualizations that allow users to compare and steer design variants [67, 74], making latent variables more tangible and steerable through direct manipulation.

2.3.4 Summary. ImaginationVellum pushes many of the above ideas forward. It takes advantage of its canvas as both output and input for richer, more expressive creative intentions to relay to its underlying generative models. It fully embraces many input channels that include location, inking, text, and images. It also uses its canvas to help users interactively explore regions of a latent space. Most importantly, ImaginationVellum opens up new possibilities of creative interactions for novice and experienced designers alike.

3 Walkthrough Scenarios

To ground the use of ImaginationVellum in real-world creative workflows, we describe two concrete walkthrough scenarios. These examples—one from an educator sketching the stage setup of a school theater production, and a second one from a maker designing a science-fiction inspired lamp—serve to contextualize the spatial prompting interaction techniques. The scenarios show creative workflows with ImaginationVellum, and foreshadow the spatially-aware canvas interaction techniques with generative AI covered in more detail later in the paper.

Scenario 1: Teacher planning theater stage setup*. For the elementary school theater production "Dinosaur Dreamland", Leon needs to explore ideas for the stage setup. He takes a photo of the stage in the assembly hall, drags it onto the ImaginationVellum canvas, and converts it into a sketch (Figure 2a). He then adds different tags to the canvas, describing ideas for the stage setup, and uses multiple Generative Strokes to see many variations at once (Figure 2b). Along the path of the Generative Strokes that Leon draws across the canvas page, ImaginationVellum creates many variations of inspiring sketches in response to the intent tags he placed on the canvas. He finds several great visual ideas for the dinosaur-inspired stage design, crops the images to the most interesting elements, layers them on top of the stage sketch, recomposes them, and creates a new combined sketch (Figure 2c). He continues exploring details of the stage design—generates, selects, resizes, and rearranges elements—and creates a final drawing (Figure 2d). Happy with the stage setup, he sends the canvas to his colleague, who curiously goes over the many creative ideas laid out on the canvas, and suggests further ideas based on sketches from different branches of the Generative Strokes.*

Scenario 2: Maker envisioning a science-fiction inspired lamp: *Xia's uncle loves science-fiction movies and she came up with the idea to craft a scifi-inspired lamp made out of metal in her maker workshop as a gift for him. Not really sure where to start—but mindful of her limited manual-sketching expertise—she types "scifi lamp" onto the ImaginationVellum canvas and scribbles a few quick pen strokes to indicate the shape of her lamp (Figure 2e). She selects the ink strokes and ImaginationVellum returns various sketched ideas based on her wavy scribbles. Intrigued by the visual results of creatively shaped lamps, and delighted by how well these sketches created by generative AI turned her own ink strokes into inspiring design ideas, she feels empowered and continues scribbling across the page—different shapes and forms she likes, in a playful back and forth with the system. The canvas page fills with designs, and to further vary the output she uses Generative Strokes, laying out variations along the path and alternating the ways her initial ink strokes get converted into design sketches (Figure 2f). She continues creating branches of variations with further Generative Strokes, taking different previously generated lamp designs as source material for a new set of variations. After some more experimentation, she finds her favorite two designs and changes sketching styles to see higher-fidelity versions (Figure 2g)—and also two technical construction line drawings that will help turn her vision of the science-fiction inspired lamp for her uncle into reality at her maker workshop (Figure 2h).*

Both scenarios demonstrate how ImaginationVellum supports the visual ideation process, with generating sketches guided by the user, quickly sampling variations with Generative Strokes, alternating between preserving and transforming sketching styles, and affording rich composition and re-mixing of content elements to control each stage of the ideation. In the remainder of the paper we unpack the conceptual and technical contributions of spatial prompting (in particular Generative Strokes).

4 Spatial Prompting

Informing the design of ImaginationVellum's spatially grounded interaction techniques, we define *Spatial Prompting* as a method for mapping generation parameters to spatial gradients. We introduce technical design choices for mapping spatial input changes to prompt and generation parameters (Figure 3), with a definition of spatial prompting in generative AI as follows:

> *Spatial Prompting maps an* ***n-dimensional spatial attribute*** *(such as location of an object, relative distance between items, presence in a particular area) to either the* ***constructed prompt*** *(e.g., changing the prompt text or weight of tokens) or* ***any generation/model parameter*** *(e.g., weight of ControlNet, temperature, or the classifier free guidance scale).*

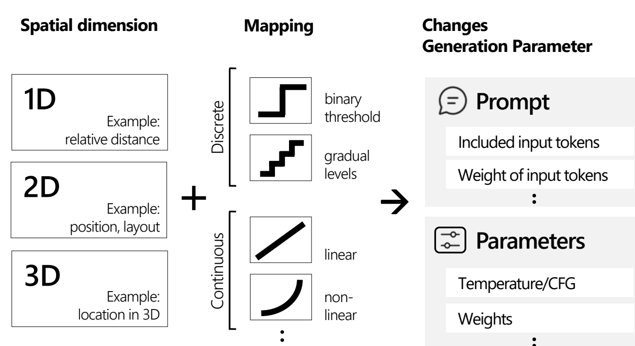

Figure 3: *Spatial prompting*: mapping different spatial dimensions (e.g., 1D changes in relative distance, or 2D layout changes) to changes of prompts (e.g., changing weight of input tokens) or other generation parameters (e.g., weights, temperature/CFG).

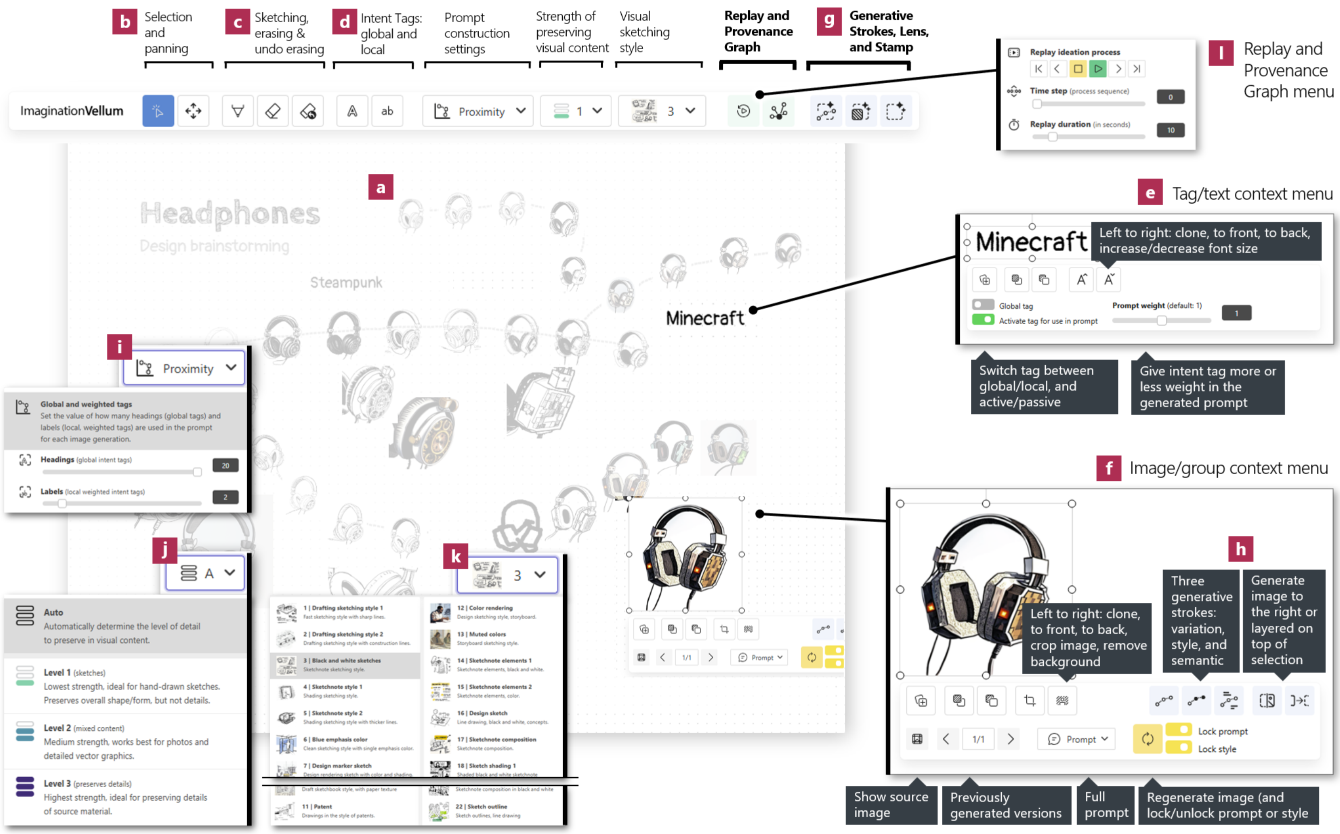

Figure 4: ImaginationVellum user interface: (a) canvas for creative co-creation with AI, adding, and re-mixing sketches, images, and text, (b–d) interface controls for selection, panning, drawing, erasing, and adding intent tags, (e) contextual control menu for text labels and (f) images, (g,h) controls for selecting generative tools such as Generative Stroke, (i–k) options and selection of visual style, and (l) controls for temporal replay and activating provenance graph.

The above definition expands on earlier work, such as World-Smith [24], where spatial prompting refers to the spatial arrangement of multiple 2D source images used to generate a combined image via inpainting/outpainting, and defining multiple regions of an image with specific prompt instructions.[1]

Technical Design Choices for Spatial Prompting. With the scope of this definition of Spatial Prompting, we can now consider different spatial dimensions and map their change in value to either changing the prompt or other model parameters. Figure 3 illustrates an overview of these choices:

- **Spatial Dimensions (Figure 3 left)** are the changes in spatial input. Examples are the relative distance between items, changes of a 2D layout, or updated 3D position.
- **Changes in Generation Parameter (Figure 3 right)** are the dimensions that are controlled and changed. Examples are a change to the prompt input tokens or weight of tokens, or a change to generation parameters such as temperature or weights.
- **Mapping (Figure 3 middle)** is the link in between the spatial dimensions and the generation parameters. Examples are a linear mapping, a discrete binary threshold where a change is happening, or gradual discrete levels. The choice of this mapping transfer function determines how generation parameters are then modulated based on changes in any of the spatial input dimensions.

This list of technical considerations is not exhaustive, we can add other parameters (e.g., top-p, penalties, or different models) or mappings (e.g., piecewise defined function). With ImaginationVellum, we applied this notion of Spatial Prompting to inform the design of tightly-coupled co-ideation interaction techniques with AI on a digital canvas. We introduce these techniques—in particular, Generative Strokes—in the following section, and we refer back to this definition and technical choices for spatial prompting.

5 ImaginationVellum

In this section, we present ImaginationVellum's design and interaction techniques. We begin by describing the design goals that guided its conceptualization, then detail the main components of

[1] Different to our definition, in robotics research *spatial prompting* refers to spatial directions given as instructions to a robot (e.g., through multi-modal gaze or pointing input) or to subtle cues influencing the spatial position of a user [81].

the user interface and freeform canvas. We show how these elements support steering the co-creation with generative AI through Spatial Prompting. In particular, we introduce Generative Strokes as a class of techniques that modulate AI generation parameters or prompts along freeform strokes. And we explain how Temporal Replay and Provenance Graphs make ideation trajectories actionable and turn the design process itself into an artifact that affords reflection.

5.1 Design Goals

Grounded in prior work on generative AI for creative ideation (Section 2), we identify four primary design goals that informed the development of ImaginationVellum:

DG1 Support freeform composition and remixing on an open canvas. A key goal is to treat the 2D canvas as an unbounded creative surface where sketches, text, and images—including AI-generated content—can be freely composed and reused. With this approach, the *canvas with all its content becomes the prompt*. Flexible layout and serendipitous remixing are central to real-world creative practice [17, 22, 37, 102].

DG2 Enable light-weight exploration, juxtaposition, and parallel ideation. Early-stage ideation benefits from rapid divergence, convergence, and comparison [33, 34, 88, 92]. ImaginationVellum aims to lower the overhead of branching and comparing ideas in parallel, helping designers swiftly iterate and visually juxtapose alternatives.

DG3 Preserve the ideation process as a visible artifact. Reflection on design decisions, alternative paths, and branching becomes possible when the process itself is explicitly recorded and made visible [35, 102]. We aim to capture and surface decisions, branching points, and evolving sketches or prompts, so users can revisit and reflect on their trajectories or share them with collaborators.

DG4 Support fluid transitions between AI generations and people's manual creative work. Creators benefit from moving between AI-generated assistance and manual expression. ImaginationVellum strives to preserve fluid transitions between automated generation and hands-on refinement or composition, so that both can be combined without disrupting creative flow.

5.2 Interaction with the ImaginationVellum Canvas & Content

5.2.1 Canvas Interface. ImaginationVellum's primary workspace is a panning and zooming 2D canvas (Figure 4a), where a person can freely place sketches, images, text boxes, or AI-generated content (DG1). The *selection tool* (Figure 4b) manipulates one or more elements, and any selected items can serve as inputs for the generative AI. A user can add content in several ways:

- **Digital pen tool** (Figure 4c): Creates pressure-sensitive ink strokes for freeform drawing or written annotations; an eraser and undo-erase mechanism allow non-destructive removal of strokes or parts of images, enabled by masks to hide content.
- **Text boxes** (Figure 4d): Provide typed text for notes or *global* and *local* intent tags (Section 5.2.5).
- **Images** (in PNG, JPG, or SVG format) can be imported, dragged directly onto the canvas, or pasted from the system clipboard.

All items can be freely resized, moved, and layered. For any selected element (or group of elements), a context menu with further options and controls is visible below the element, as shown for the text box in Figure 4e and the image in 4f. A user can select generative tools either from the menu bar (4g) or the context menu of an image or group (4h), in which case the selected element or group forms part of the input for the generated content (e.g., transforming the selection into a new sketch). Generation parameters can be fine-tuned in various panels: prompt construction (4i), strength of preserving visual input during generation (4j), and selected visual sketching style (4k). The top menu also offers features for temporal replay and provenance graph (Figure 4l).

5.2.2 Free-form vs. Structured vs. Hybrid Canvas. There are different strategies for designing digital canvases. Some rely on free-form *unstructured* 2D space (e.g., [48, 77, 82, 103]), while others adopt *structured* layouts like node-link diagrams [18, 42] or spreadsheet-like grids [4]. ImaginationVellum uses a free-form canvas for the ideation process, as this strategy is more closely aligned with enabling creative workflows and rich manipulation of content. On the other hand, the system also leverages the layout and grouping relationships of images, text, and sketches as input dimensions. Therefore, our design combines both worlds, in a *hybrid canvas*, which on the one hand allows free-form manipulation and composition of elements for creative expressivity, and on the other hand uses the structure and layout of elements to steer the generation of content using spatial prompting.

5.2.3 Sketches as Scaffolding Material that Invites Participation. Sketching is an important asset in the early-stage design and ideation process. However, being a learned skill, not everyone is comfortable with the prospect of using sketching as the primary means for interaction. We implement two methods to facilitate this:

(1) First, any content in ImaginationVellum is generated in a low-fidelity sketched style. Low-fidelity sketches are a *material that invites participation*, encouraging users to co-sketch and co-ideate with the generative AI independent of their sketching skills [17, 33].
(2) Second, we augment the user's own sketching with generative AI. ImaginationVellum can take any hand-drawn sketches by the user and create generative AI driven sketches based on those scribbles (examples shown in Figure 5). This approach encourages iterative user scribbling over AI sketches, layering changes, and quickly regenerating new variations, without implying a final polished artifact.

In this way, the sketches become a material substrate, lowering the threshold for participation, and providing the right fidelity at the right time [33]. As users advance through their ideation process, they can adjust the sketching fidelity in the system, for example, using lower-fidelity ambiguous sketches during early ideation, and higher fidelity design sketches during later stages.

To accommodate a variety of visual preferences and design phases, we curated 22 reference sketch styles, loosely categorized into: (1) rough hand-drawn and drafting sketches which are ideal for

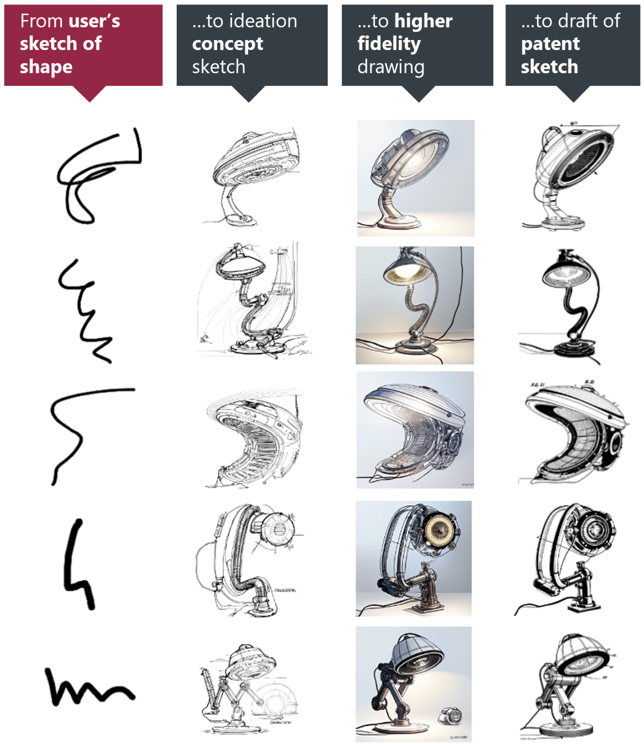

Figure 5: Examples of the different sketching styles recreated by ImaginationVellum using Stable Diffusion and Reference ControlNet (Prompt: *"scifi lamp"*, from the walkthrough scenario): from left to right, a person scribbles on the page, turns this into a lower-fidelity idea sketch, a higher fidelity drawing, and a technical drawing sketch.

early ideation, (2) sketched black and white line drawings which create sharper and more defined lines and decrease ambiguity, and (3) higher fidelity design sketches and marker drawings. The user can at any time switch styles for their own sketches, a previously generated image, or a selection of multiple objects (Figure 4k). We use a reference style ControlNet [111] to preserve each of these sketching styles when generating new content, to create a co-sketching experience that brings a user into the flow of sketching with the AI by locking the fidelity of generated images.

5.2.4 Sequences, Composition, and Re-mixing. ImaginationVellum encourages *turn-taking* between the user and the generative model during ideation. Users may import a photo, convert it into a sketch, then make freehand edits, prompting the AI to generate new versions or partial re-draws in an *iterative revision sequence* (Figure 6a). Beyond revision, users can *re-mix* existing generated artifacts: for instance, layering an image of a plant with an architectural photo to inspire a new hybrid design (Figure 6b). From the first results, the shape of the original plant does not come through in the design, so there is another composition with the images, leading to a new design of the building. Figure 6c shows an example where a user re-mixes previously generated images of a lamp design by composing different elements of images together, adding missing details by sketching, and finally initiates the generation of a new sketch. These quick manipulations treat **generative-AI images as a material** for further creation and help mitigate challenges in creative human-AI workflows like design fixation [101].

5.2.5 Steering Content Generation with Proximity-Dependent Intent Tags. To guide AI outputs, ImaginationVellum extends the concept of *micro-prompting* [32] with *local* and *global* intent tags. A user can place *intent tags* as text boxes on the canvas (Figure 4d).

- **Global Intent Tags** apply to the entire canvas regardless of their position on the canvas.
- **Local Intent Tags** vary in weight according to spatial proximity. Images generated near a local tag will be influenced by that tag more than those placed farther away.

Combining global and local tags allows the user nuanced control—maintaining a broad language while selectively experimenting with local variations (e.g., distinct shapes, textures). The next section outlines how these intent tags are used by generative tools to construct prompts for generating new content.

5.3 Generative Strokes

We designed a set of interaction techniques that leverage *the canvas as a prompt* to create new generated content. These techniques leverage spatial and semantic relationships to modulate prompt construction and guide the creation of new images (an overview of the techniques is shown in Figure 8). A user can select these generative tools directly from the main toolbar (Figure 4g) or the contextual menu of an image or group of objects (Figure 4h).

5.3.1 Semantic Generative Strokes and Prompt Construction. With *Semantic Generative Strokes* (G1), users can rapidly define where to add generative sketches by simply drawing a freehand line across the canvas. As the user draws a line, the system creates a series of generative images at intervals along the path. For each image, ImaginationVellum computes its distance to any local intent tag, as well as the global tags placed on the canvas, then constructs a prompt with the appropriate weights.

For instance, Figure 7 shows the generation of "retro chair" concepts with a generative line that transitions between "fluffy,rounded" and "blocky,sharp,spikes" intent tags—the resulting images smoothly interpolate between these concepts. The spatial prompt is constructed by dividing the *distance from the image to surrounding local tags* by the *maximum distance between the tags* (limiting to the range [0..1]). This distance factor is then mapped with a linear transfer function (see section 4) to change the weight of prompt tokens, forming intent tag weights such as "(fluffy,rounded: 0.25)" and "(blocky,sharp,spikes: 0.91)." For each image, the complete prompt is constructed as `prompt = [n x global intent tags], [m x (local intent tags: weight)], [reference style tokens]` using the following elements:

- `[n x global intent tags]`: all active global intent tags placed on the canvas are added to the prompt, up to the limit `n` (configurable in the interface: Figure 4i).
- `[m x (local intent tags:weight)]`: This adds the `m` closest local tags (by default, `m=2`) to the prompt, weighting each according to distance. In Stable Diffusion, weights > 1.0 increases emphasis, while lower weights decrease it. Through in-house

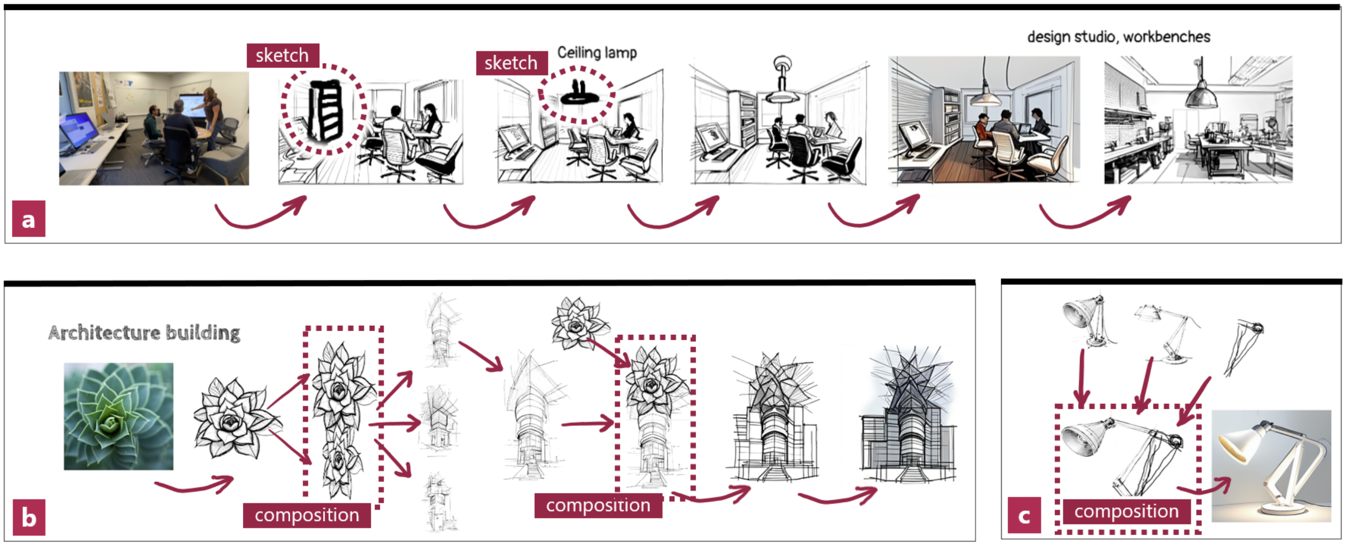

Figure 6: Examples of iterative workflow sequences and re-composition on the canvas (Actions done by the user are highlighted in red, such as sketching, composition, or requesting generated images—all other content is generated by ImaginationVellum): (a) iterative redesign of an interior space, with ImaginationVellum creating new revised sketches based on the user's drawings on top of the sketches, choices of style, or added intent tags. (b) composition of content elements is used to guide consecutively generated images, and (c) another example of re-mixing images and sketches together to generate new content.

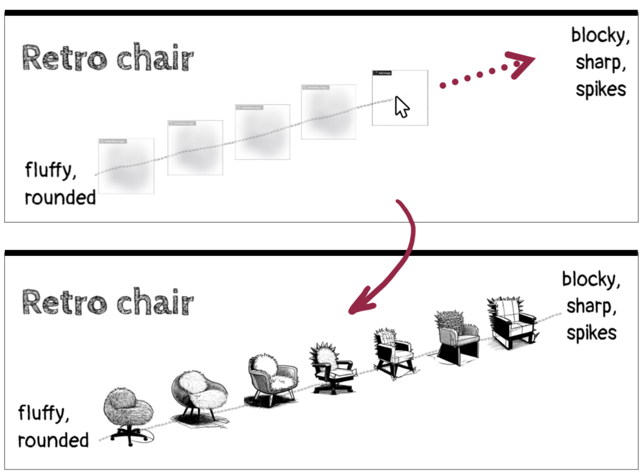

Figure 7: Example of a Semantic Generative Stroke blending content defined by two intent tags.

testing, we found that values above 1.3 can distort the generated content. Thus, we scale weights from 0 (at maximum distance) to 1.3 (at minimal distance). Additionally, we set the maximum radius for each tag to half the canvas size. Both the number of local tags to use as well as the maximum radius for each tag are configurable in the interface (Figure 4i).

- `[reference style tokens]`: These are descriptive tokens that correspond to the currently selected sketching style (e.g., low-fidelity pen sketches), complementing the visual reference image used to control generation.

Two related tools, *Generative Stamps* and *Generative Lenses*, provide further ways to generate new content on the canvas. *Generative Stamps* place a single image at the cursor location, mirroring the logic of Generative Strokes but without continuous interpolation. This tool is useful for quickly placing images across the canvas. Building on the notion of Transformative Lenses in [82], *Generative Lenses* transform the content beneath them through a generative process guided by nearby intent tags. They act like filters that remix the underlying content, based on semantics of nearby intent tags, while retaining the spatial context.

5.3.2 Rapid Exploration of Design Spaces. Generative Strokes afford rapid exploration of design spaces by allowing the user to scatter multiple global and local intent tags across the canvas, then drawing strokes that blend or recombine these concepts.. Figure 9 shows two examples. In (a), we place three local tags to create a triangular design space similar to Scott McCloud's visual iconography [62]. Images along the edges interpolate between pairs of tags; images in the center blend all three. Critically, these tags do not form a fixed structure—they can be arranged flexibly on the canvas. As the resulting empty spaces between them is filled using our generative tools, it affords the flexible and creative scanning of a multi-parameter design space. The second example, Figure 9b, demonstrates using the Generative Stroke across five local tags to generate boathouse designs that blend the semantics of each tag.

5.3.3 Class of Different Generative Stroke Tools. While Semantic Generative Strokes (G1) focus on blending local and global intent tags, the same principle of modulation in response to spatial gradients can be applied to other parameters embedded in the canvas. We

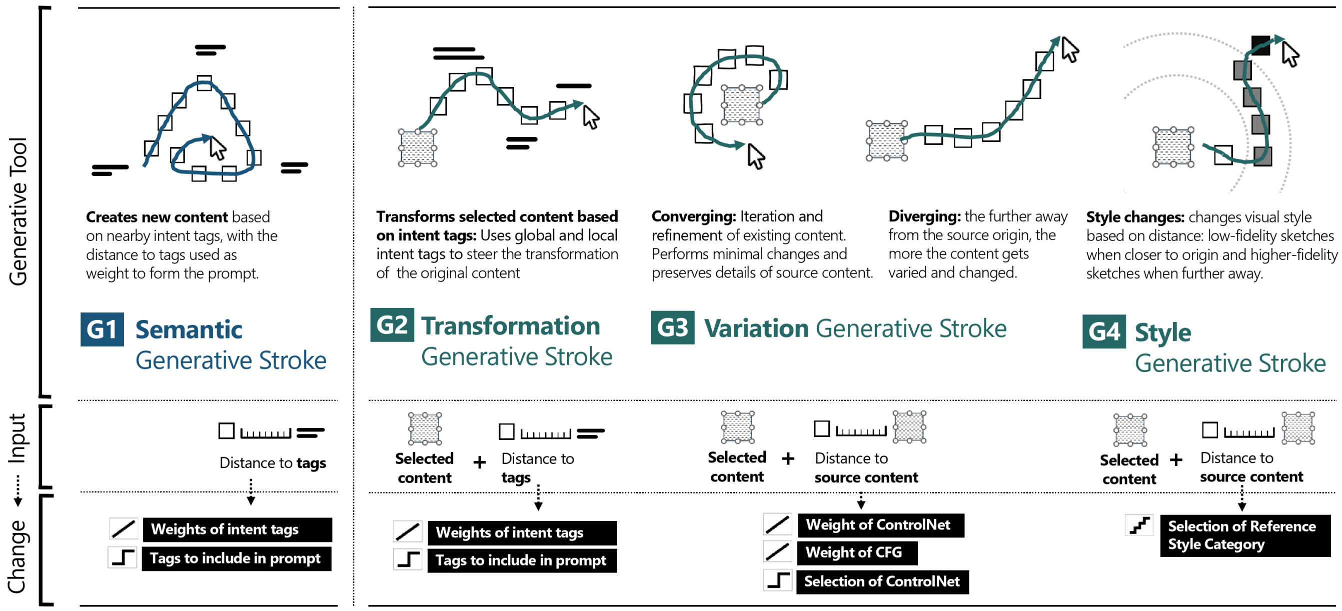

Figure 8: Overview of different classes of Generative Strokes: (G1) Semantic Generative Stroke, creating new content based on spatially proximate intent tags, (G2) Transformation Generative Stroke, transforming selected content based on semantics of nearby intent tags, (G3) Variation Generative Stroke, which based on distance creates either converging or diverging alternatives of the selected content, and (G4) Style Generative Stroke, which changes fidelity of visual sketching style based on distance.

designed three additional Generative Strokes (G2–G4) that transform selected content (images, inked sketches, or any composition) by controlling different generation parameters along the generative stroke (see Figure 8).

- **G2 | Transformation Generative Stroke** transforms a selected image or sketch guided by spatially proximate intent tags. Prompt construction remains similar, but instead affects only the visual appearance. We use a "Scribble/Sketch" ControlNet layer to preserve details of the selected content source.
- **G3 | Variation Generative Stroke** supports both convergent and divergent interactions over a design, depending on how a user draws the stroke relative to the source. As shown in Figure 8-G3, drawing close or looping tightly around the selected object yields only subtle modifications, whereas dragging the stroke away from the object leads to more divergent designs. We achieve this by mapping the distance to three key factors: (1) the chosen ControlNet (Canny Edge when closer for fine variations, and Scribble/Sketch when further away for larger changes), (2) the ControlNet weight [0..1], and (3) the classifier free guidance scale (CFG) [4.. 10]. Both the weight and CFG are higher when close, smaller when further away.
- **G4 | Style Generative Stroke** uses three distance zones around the selected object to determine a sketching style: the closest zone yields low-fidelity drafts, the medium zone create medium-fidelity line drawing sketches, and the furthest zone switches to high-fidelity illustrations, allowing the user to easily explore different sketching fidelities.

All these Generative Stroke techniques illustrate how the distance along a freeform path can modulate latent or stylistic parameters, such as semantic interpolation between tags, transformation of selected content, convergent iteration and refinement, divergent idea generation, or switching between low- and high-fidelity representations. Some of these mappings are linear (e.g., tag weights in G1 and G2), others stepwise or discrete thresholds (e.g., binary switch between ControlNets in G3 and graduated levels for reference styles in G4). In future work, this design space of Generative Strokes can be extended to many other latent dimensions.

5.3.4 Override and Manual Control. While the proximity-dependent techniques are expressive tools to steer generation, ImaginationVellum provides options to override this behavior and use manual control for prompting and content generation. For instance, users can decide to exclusively use global tags, which are not weighted depending on proximity when added to the prompt. Used this way, a single global tag becomes a direct prompt input field, similar to most image generation websites. Advanced users can also use stable diffusion prompt markup, such as round and square brackets to increase/decrease the weight of terms directly.

5.3.5 Feedforward of Spatial Prompt Input. Feedforward techniques in interfaces can provide a user advanced information and preview of an action's effect [100], and previous research indicated that such previews of AI outputs before prompt submission can aid user control [64][2]. In ImaginationVellum, whenever a user invokes a generative tool—such as Generative Strokes—feedforward

[2]The use of the term 'feedforward' in interaction design is different to the terminology in machine learning, where feedforward indicates information flow in neural networks.

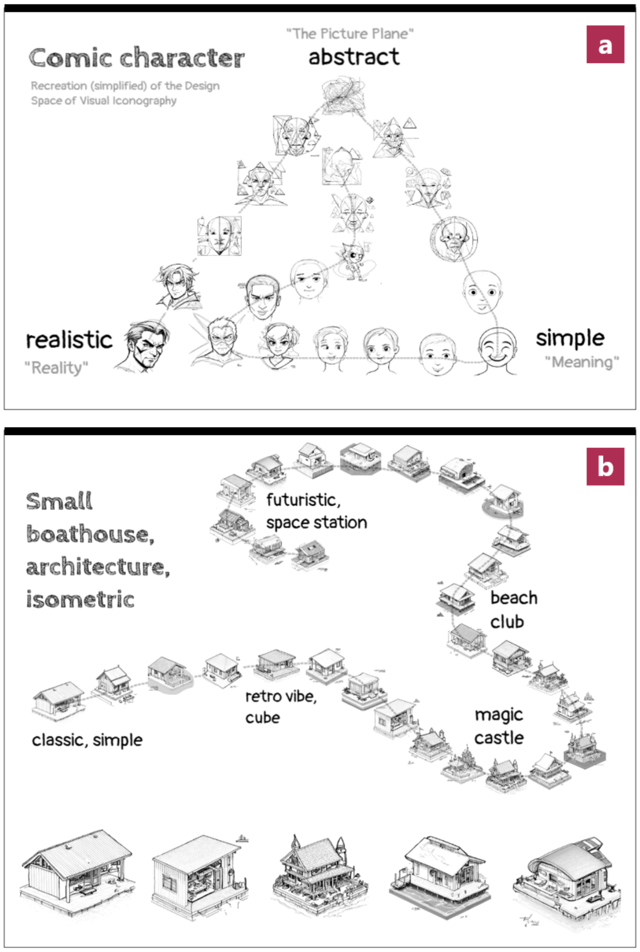

Figure 9: Examples of using Generative Strokes to explore design spaces along a path the user draws on the canvas: (a) Generative Stroke creates blended variations based on local intent tags as endpoints (a simplified version of McCloud’s iconography design space [62]), (b) Generative Stroke along five local intent tags to quickly sample different designs of a boathouse.

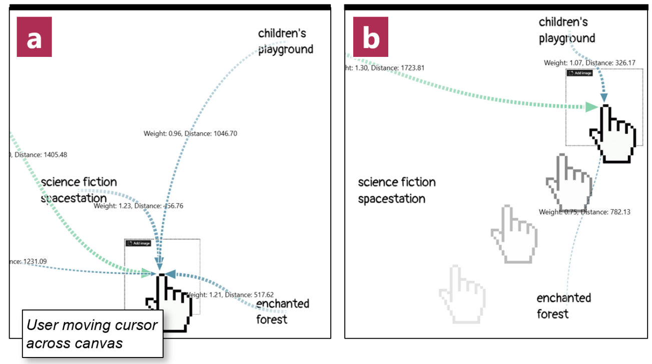

Figure 10: Feedforward: (a,b) arrows indicate local and global intent tags that will be part of the prompt if the user decides to generate content at the current cursor position (line weight is modulated by weight of the prompt).

cues graphically indicate how nearby content elements on the canvas will influence the generated outcome. Once a generative tool is selected in the interface, arrows extending from nearby intent tags extend to the user’s moving mouse cursor (Figure 10). This includes arrows from all global and local intent tags that would be part of the prompt for any content generated at that location on the canvas. The thickness of lines visually encodes the weight of local intent tags. This dynamic visual cue updates as the cursor moves, ensuring the user sees how positional changes alter the prompt—similar to OctoPocus’s “dynamic guides” [9]. While real-time preview feedforward techniques often pose computational challenges [25], an alternative feedforward technique in ImaginationVellum could show partial or approximate live previews of generated content at the cursor position.

5.4 Revealing Ideation Process: Temporal Replay and Provenance Graph

One of our design goals (DG3) is to support reflection on the design and ideation process. By design, ImaginationVellum puts emphasis on preserving the journey of the ideation process on the canvas page, by using tools that lay out alternatives, allow branching, or re-combine and re-mix content—and by keeping this content and variations on the canvas itself. We argue for the *design process as an artifact itself*, emphasizing the value of making the creative or design process visible for effective collaboration and sensemaking. We encourage both introspection and sharing of creative journeys, including the whole canvas, with others—similar to how people might share their physical ideation sketches with collaborators. We implemented two techniques to make the ideation process more visible:

(1) **Temporal Replay** can play back the timeline of canvas edits. Every action on the canvas (e.g., content created by the user or generative AI) is time-indexed, allowing replay of the evolving ideation canvas in a sped-up review (for example, as a 10 second summary). Unlike a simple video, this replay is fully interactive: the user can pause at any point and branch the design from that moment, creating a new canvas page that inherits prior elements. This facilitates revisiting earlier ideas or exploring alternative directions without discarding existing work.
(2) **Provenance Graph** reveals the provenance of an element. Upon hovering over an item, this highlights all elements (e.g., sketches, generated images, compositions, and intent tags) that contributed to it, layering them in a color-coded tree structure (Figure 11a). Items outside the inheritance chain fade to 20% opacity, providing clarity without clutter. Figure 11b-d show other examples of canvas pages, and an example provenance graph for selected items. Critically, the Provenance Graph only appears on demand to minimize visual noise. This offers a compromise between a minimally structured freeform canvas for exploration, while also supporting detailed revelation of the underlying ideation journey on demand.

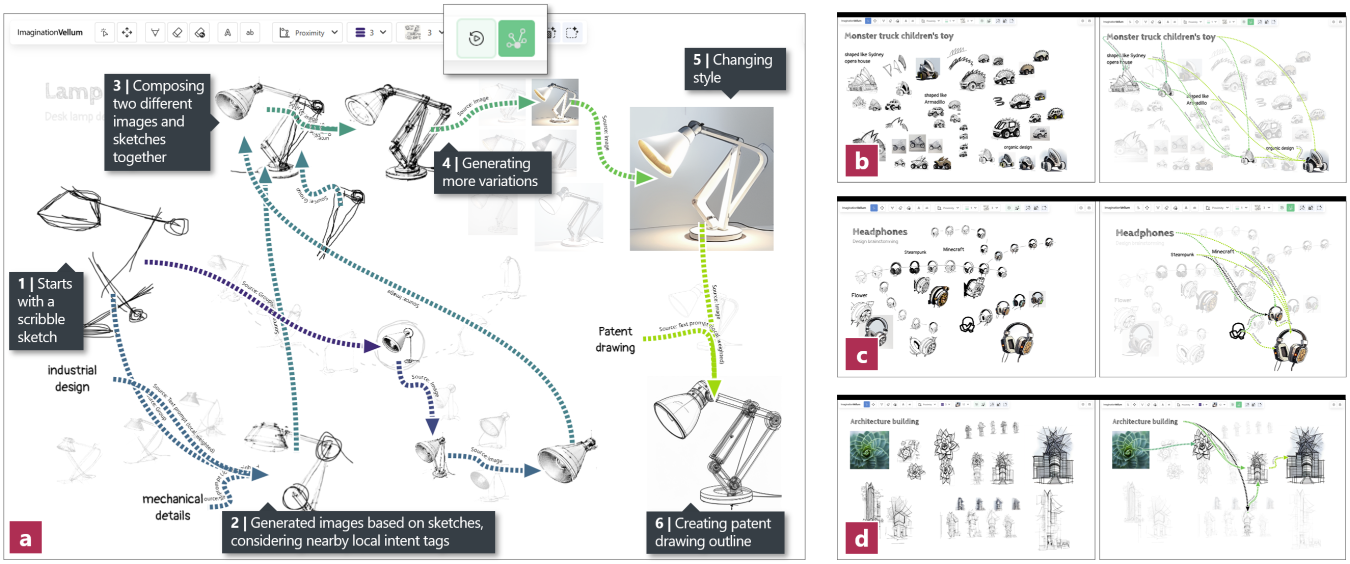

Figure 11: Reviewing the ideation design process on the canvas: (a) provenance graph reveals how intent tags, sketches, previously generated content, and compositions of artifacts have influenced the patent drawing outline in the lower right corner. (b–d) Other examples of provenance graphs.

5.5 Implementation

ImaginationVellum is implemented as a web-based architecture (Figure 12), with JavaScript for the front-end, and REST API services for the back-end, processing, access to LLMs/VLMs, and Stable Diffusion.

The front-end uses JavaScript/HTML and Fabric.js [44] library, and communicates with the back-end Node.js [31] server using Express.js [30], which synchronizes the access to multiple Stable Diffusion [94] instances. We use a locally hosted Stable Diffusion [94] server (powered by an NVIDIA RTX 4090 GPU) with a custom processing pipeline for image generation. We also use the Phi-4 [1] Large Language Model (LLM) for text analysis and prompt augmentations, and Llava-v1.5-7b [56] Vision Language Model (VLM) for image/sketch analysis and handwriting recognition.

For image generation, we use multiple ControlNet [111] models with Stable Diffusion to steer the generation of visual outputs (illustrated in Figure 12). To preserve content of the source input, we use a selection of *Depth*, *Canny Edge*, and *Scribble* ControlNet models (choice of the model depends on how much detail of the original content to preserve). For constraining outputs to sketching art styles we use the *Reference Style* ControlNet model. The reference module takes one of 22 visual sketching style images (created by the co-authors of this paper) as reference sources to recreate the characteristics of each of these sketching styles. We iteratively refined additional prompt descriptions for each style, which get added to each image generation request (as positive and negative prompts), in addition to the prompt constructed from all intent tags on the spatial canvas. Depending on the type of image generation, we vary the weight of each ControlNet model (e.g., as mentioned earlier with the Generative Stroke techniques, where we change weights, prompts, and other parameters to emphasize or de-emphasize content preservation). We use image masks and filters for image manipulations, such as increasing contrast or turning sketches into transparent PNG images. ImaginationVellum modifies other parameters such as CFG scale, denoising strength, and control mode based on the selected interaction technique and the spatial prompting parameters.

6 Qualitative User Study

We conducted a qualitative user study with six participants to gather preliminary feedback from both novices and professionals about the concepts and techniques within ImaginationVellum. The study was exploratory and small-scale, designed to surface early reactions to ImaginationVellum's interaction techniques—in particular Spatial Prompting and Generative Strokes—and we report on participants reflections on expressive control through spatial prompting, revisiting the co-creation process, and differences in values for creative workflows with generative AI.

6.1 Study Design

Procedure. The experimenter introduced the study structure soliciting genuine feedback on both positive and negative aspects of the novel techniques of sketching and ideation. After signing a consent form, indicating their agreement for audio and video recording, participants sat in front of a Microsoft Surface Studio (28-inch 4.5K display), inclined as drafting table mode, and had a digital pen available as well as a mouse and keyboard. The experimenter sat next to them and used a guided walkthrough methodology, providing step-by-step instructions for participants to execute, explaining underlying principles of each sets of technique (e.g. proximity). The guided walkthrough consisted of three parts:

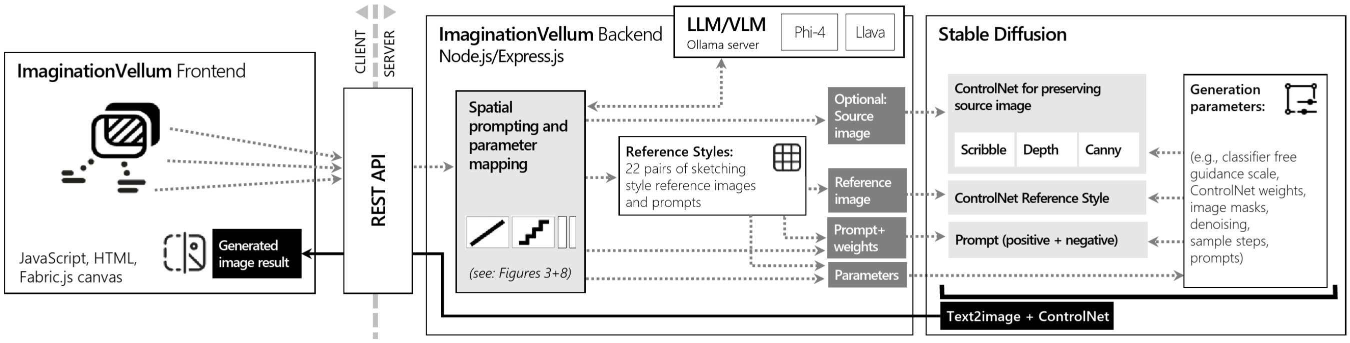

Figure 12: ImaginationVellum architecture and implementation.

(1) using generative stamps with labels, sketches, photo import, and interactions available for composition (crop, transparency, inking, generating new sketch from selection);
(2) canvas generation replay and provenance graph on hover; and
(3) generative stroke techniques.

After each part, participants had five minutes to experiment on their own, before the experimenter collected their initial thoughts and reactions. The experimenter asked about their sense of control of the generation, the cognitive ownership of the resulting sketches, the perceived value of the techniques for divergent and convergent thinking.

After the guided walkthrough, which lasted 30 to 45 minutes on average, the experimenter asked participants to freely use ImaginationVellum for a task of their choosing. The experimenter reserved the final five minutes of the session to gather participants' impressions on most salient strengths and weaknesses of the experience, as well as their thoughts for future directions.

Participants. We recruited 6 participants (3 men, 3 women) via mailing lists in a large organization. We selected participants with prior experience of generative AI systems (ChatGPT, Stable Diffusion, etc.) for creating visual content. The background and skill set of the participants can be found in Table 1. The study was approved through the internal Microsoft Research Ethics Review Program. We conducted one hour in-person study sessions and participants were compensated with $75 USD gift cards.

Data Collection and Analysis. We recorded each study session on video and transcribed the audio. We also recorded the screen, and used features in ImaginationVellum to save each created canvas, including the provenance graph and detailed timeline of events. We inductively coded and analyzed all transcript for emerging themes, and reviewed the saved canvas files, screen recordings, visual content, and timelines of used tools/functions.

6.2 Findings

Canvas interaction to steer visual content generation gives user a higher sense of control than using text input alone. All participants commented on the value of steering image generation using sketches *"I love that you could draw something and then create something from that"* (P3). In particular, they commented on having a greater sense of control in contrast to using text input: *"being able to have more control over the sketch versus just letting the AI decide it for you"* (P5).

Dragging images as references was also unanimously praised by our participants. P6 articulated the value in contrast to prompting: *"With text [prompting] is so it's so hard to actually get [GenAI models] to do something interesting, whereas here it's like, oh, I can show you with images. Like I really wish you had more skyscrapers here, or I really wish there was a car. And then you can just kind of pop that in there and say, you know, try it again. It's just a different way of communicating with it."*

Canvas as an affordance in contrast to just text prompting. P2 noted that a non-linear canvas was more aligned to their design process noting that it was *"way more naturally aligned to my cognitive style and preference for doing anything with visuals."*. They hinted at a different cognitive mechanisms when crafting prompts as opposed to ImaginationVellum matching the natural way they worked on paper: *"the biggest Pro is not having to get out of my visual flow and be in a screen where I'm forced my brain is forced or organize differently. Even though my brain can organize fine around words, it's not necessarily enjoying that when I'm doing something more creative or generative per se. So the biggest thing is that I don't have to put in prompts."*

P6 commented that ImaginationVellum and prompting provided very **different affordances**: *"I think that when a canvas is in front of me [...] it puts me in the mode of, like, OK, I want to draw a thing and have it follow that as opposed to sometimes I'll approach a text image generation situation where I'm kind of like I have no idea what a leprechaun who's eating a brownie would look like, and it's kind of like just make one for me and it's like, oh, yeah, that's pretty good."*

For this reason, P6 felt that when the technique failed to capture their intent behind their sketch, it was more frustrating than direct prompting. *"with text I might be more forgiving because it's like I don't have an idea, I just type something."* Furthermore, P6 suggested having an option to *"bring up a context-free prompt window"* (P6) separate from the canvas—essentially a blank prompt box—for situations where one just wants to type a quick idea without the context of the spatial canvas, and then import the resulting image into the canvas to integrate or remix. This feedback highlights that while the spatial canvas workflow was generally seen as powerful

	Participant	Background	Sketching Skills	Design Workflow and Generative AI Tool Use
NOVICE				
	P1	Management	limited	–
	P5	Executive Assistant	limited	–
PROFESSIONAL				
	P2	Design Researcher	experienced, pen and paper	Figma, Dreamstudio
	P3	Professional Designer	experienced, and sketches in free time	Figma, Adobe Illustrator
	P4	Professional Maker	experienced, daily, pencil and paper	3D modeling, Maya
	P6	Artist	experienced	Procreate, Procreate Dreams

Table 1: Background, sketching skills, tools part of current design workflow, and generative AI tools used by study participants.

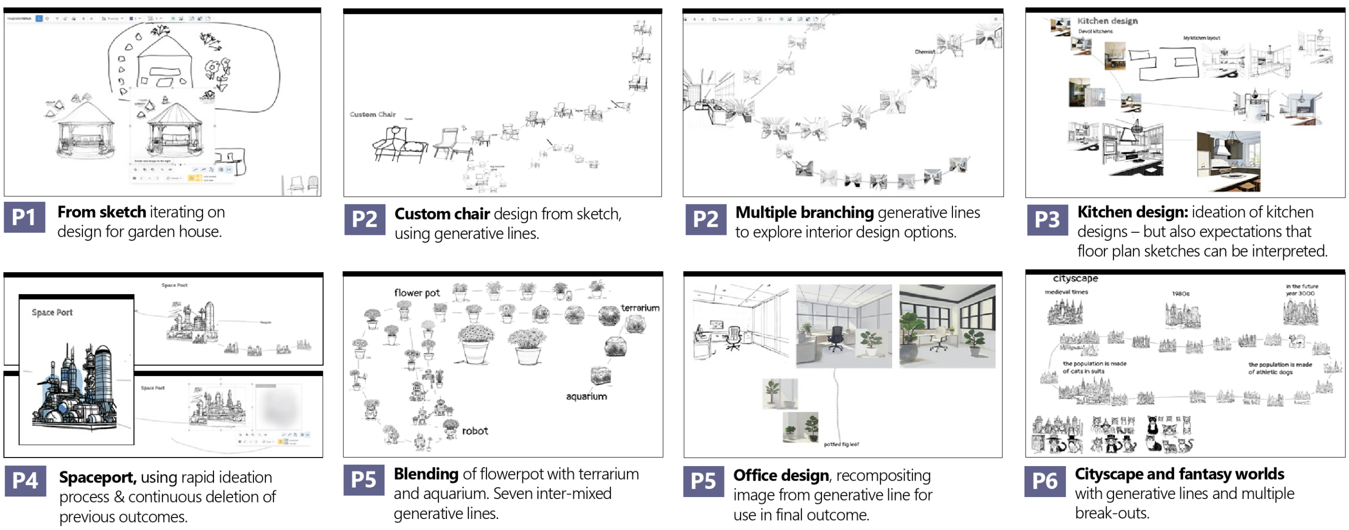

Figure 13: Visual ideation with ImaginationVellum, created by participants during our user study. These explorations were done after study participants completed the walkthrough, and participants created content based on their own interests and the topics they wanted to explore.

and expressive, users might still desire a way to use simple isolated one-off prompts (especially for very divergent ideas) and later incorporate those results into the spatial workspace.

Revealing both the human and the AI creation process with provenance graphs and replay. P2 and P6 praised the value of replay and provenance graphs to recall their creative process (examples shown in Figure 14). P2 commented: *"I love it because I get in the zone and I cannot remember... I have a feeling I would do a lot with that."* while P6 explained that months after a digital art creation, they often struggled to remember how they arrived at a particular effect or result.

P3 qualified the provenance graph of *"extremely useful"*. They explained how it revealed what went into the AI generation process *"takes some of the mystery of what the generative creation is doing"*, and suggested to interact with the links of this graph to adjust weight of factors influencing the generation of the artifact.

P4 was very enthusiastic about the replay capability and explained this was matching their own process with sketchbooks: *"It's always important for me to go back and look at sketchbooks from months ago [...] because I was in this type of emotional state and [...] I changed so I can actually derive new ideas because I'm in a different frame of mind, I'm in a different emotional state"*. They saw value in using replay to branch from a prior step: *"Go back and then start from that point. [...] I'm going to diverge. I'm going to go back to step five and then start a new branch and then that new branch, that new train of thought it matches closer to what I'm feeling at that moment"*.

Different value for different participants. P1, who has limited sketching skills, saw value in ImaginationVellum to **enhance her drawing skills**. They commented *"I love that it's able to enhance my skills because I have none."* and after a few interactions was very excited about the results *"This is fun. [experimenter: So why is this fun?] Because I can't draw and it's making things for me."*.

P5, who has limited sketching skills as well, highlighted the value of such canvas to bridge the gap between people with different drawing skills, allowing them to **create together**: *"I totally see doing that with my daughter as a quality time. She is a really good artist, so she and I could like spending hours like playing with this. I think it can bring people together, you know."*

P4, who has advanced sketching skills, saw value in ImaginationVellum for **jump starting his creation process**. They reflected on generative AI noting that existing models provide a final output rapidly and that they were not interested in that: *"I don't want it to replace my drawing, I want it to help me conceptualize."* After the study they concluded that an intelligent canvas like ImaginationVellum would not replace them (*"I am the artist"*) and instead help jump start his creation process with rapid ideation *"This is dope. This is mega dope. This is wow. This is in terms of just generating content for ideas to riff in order to grow new ideas [...] I can see a lot of strength in how I can utilize this like I use a sketchbook, except I'm shortening the amount of time that I'm putting pen to paper so I can get there [to a starting point for a maker/art project] much quicker."*

P2, a professional designer researcher, saw value in ImaginationVellum for **supporting her creative workflow** for both convergent and divergent thinking. They noted *"I would completely use that both ways... it's convergent to illustrate a point... but for divergent thinking too, because it does do some things you're not expecting it to do."*

P6, a digital artist, saw value in ImaginationVellum for **convergent thinking** *"a frustration I have with [text to image models] it's kind of like you get something back and it's kind of like, but I want to shift it in this way. So kind of doing that iteration through little clips of images [...] this is exciting to me"*. On the other hand, P3 highlighted the value of generative line techniques for experimentation and **divergent thinking** *"great for creating explorations and to see a lot of different things"*.

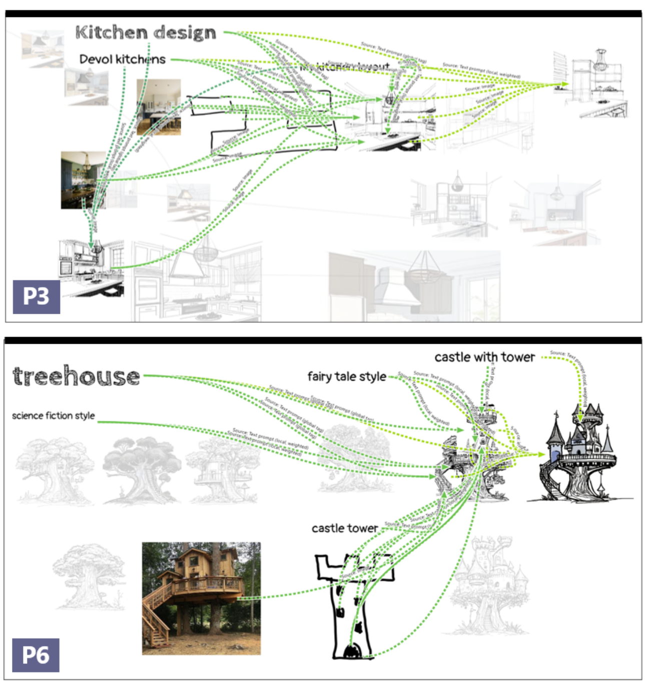

Figure 14: Provenance graphs used by participants while reflecting on their own created canvas (screen capture from study recordings of P3 and P6).

7 Discussion

Here we discuss several research issues raised by our implementation, interaction techniques, and user's reactions to them, including spatial prompting, the ideation process itself as a design artifact, control vs. serendipity in generation, and support for divergent vs. convergent thinking.

Spatial Prompting as Interaction Paradigm. Our core contribution extends traditional prompt-based AI interactions into the two-dimensional spatial domain by introducing the high degree of inputs a canvas affords (structure, images, writing, sketching, etc.) as elements to build rich expressions of intent. By mapping spatial attributes—such as proximity and layout—to prompt parameters and generation controls, our work positions the canvas as an active medium in *generating* creative outputs, allowing the 2D canvas to embody semantic meaning beyond organization through spatial arrangement. This spatial prompting approach builds on prior work that explored region-based constraints [24] and continuous modulation of generation parameters [82]. While one could argue that the spatial metaphor as currently realized in ImaginationVellum may not scale well to large-scale or more complex problems, at least for the case of early-stage ideation of individual design concepts, our results suggest that spatial prompt variations can effectively mediate user intent, thereby providing a more expressive mechanism for steering AI outputs compared to conventional text-based prompts [13, 95, 109]. Taken together, ImaginationVellum's design decisions, implementation choices, and interaction techniques illustrate practical approaches to realize our high-level design vision where *the canvas is the prompt.*

Process as Design Artifact. Another important aspect of our system is the emphasis on process preservation. ImaginationVellum integrates timeline replay and provenance graphs to document the evolution of ideas, treating the ideation process as an artifact for reflection. This capability supports self-assessment, iterative refinement, and opens the opportunity for collaborative discourse on design decisions. Through detailed capture of user interactions, the system facilitates retrospective analysis and knowledge sharing beyond conventional product evaluation. This process-level documentation proves especially beneficial for iterative workflows, where revisiting previous states can help novel design directions or facilitate branching explorations [32, 55].

Control vs. Serendipity—Balancing Acts in the Creative Process. One of the tensions we identify in our work is between user control and serendipity. Spatial prompting tools, such as proximity-dependent intent tags and generative strokes, empower users by providing fine-grained control over the generative process. By allowing users to specify weights and parameters through spatial arrangements, the system creates a tangible sense of agency over the evolution of an idea. However, this control in the *process* is complemented by the serendipity of the *output* artifact. These "happy accidents", while an inherent part of using a probabilistic model, are also essential drivers of creativity. The unpredictability of the AI output, then, becomes a resource for innovation, steered by the user.

Divergent vs. Convergent Thinking. Tools supporting the design and ideation workflows often delineate divergence and convergence as discrete phases–mirroring the rigid double diamond framework–where one is expected to shift cleanly from ideation to refinement. However, such categorical transitions rarely capture the reality of creative practice. Instead, our approach embeds both divergent and convergent modes into the materiality of the tools themselves. ImaginationVellum is purposefully designed to accommodate the rapid generation of alternatives (divergence) alongside mechanisms for selection and refinement (convergence). For instance, techniques like generative strokes promote divergent thinking by enabling rapid sampling of numerous possibilities, which can later be iterated upon and refined. This inherent duality enables users to fluidly navigate between exploring a latent idea space and focused synthesis, supporting an iterative process where serendipitous discovery coexists with intentional refinement. ImaginationVellum frames the notion of creative progression not as a binary state but as a continuous spectrum, reflecting the messy, non-linear nature of creative ideation.

Trade-offs and Implications for Creativity. Boden distinguishes creative acts that *transform* a conceptual space (*h-creativity*) from those that *explore* an existing one (*p-creativity*) [12]. In ImaginationVellum, users realize *transformation* when they lay out intent tags and compose new spatial relations, effectively re-shaping the underlying idea space. Conversely, Generative Strokes automate rapid *exploration* by populating that space with low-friction variations, letting users sample regions.

While Mixed-Initiative Creative Interfaces (like ImaginationVellum) can improve human creativity [53], using generative AI in creativity tools still carries risks. Fixation on early visual outcomes or suggested design pathways may anchor the user prematurely, limiting further exploration [101]. Further, novices often accept system defaults (e.g., with system prompts), giving up creative agency [109]. Finally, practitioners using the same foundation model can steer designers towards similar aesthetic directions, threatening stylistic diversity at the field level [75].

In order to mitigate some of these risks, GenAI-augmented creative interfaces should therefore expose contestable AI decisions via editable micro-prompts, adjustable tag weights, and provenance trails. Additionally, tools should support model pluralism through quick model swap, domain-specific fine-tuning, and user-contributed style inventories to counteract monoculture of aesthetic styles. Finally, treating every temporal snapshot as a re-entry point can encourage deliberate departures from dominant solution paths and enable reflexive branching. ImaginationVellum embodies several of these principles: users can rewrite tag text, drag weight sliders, inspect provenance trails, switch sketch styles, and rewind the canvas to fork alternative directions. Future longitudinal deployments should examine whether such mechanisms merely mitigate convergence, or genuinely cultivate new conceptual spaces—i.e., whether p-creativity ultimately ladders up to h-creativity.

8 Conclusion and Future Work

ImaginationVellum introduces a new approach to early-stage visual design ideation, by reimagining *"the canvas as the prompt"* through a user experience that engenders an expressive space for co-creation with generative AI in compelling and nuanced ways. Via spatial prompting, proximity-dependent intent tags, and Generative Strokes, the system unlocks articulate and expressive outlets to create, traverse, remix, and refine visual ideas. At the same time, it also makes the ideation history visible, explorable, and reusable—turning the often ephemeral ideation thoughts and decisions into a tangible artifact.

This work opens a promising design space for spatial prompting interfaces for fluid traversal of idea spaces with generative AI. Yet, many open research directions remain. What new creative workflows emerge when rich spatial layouts are treated as living prompts? Which other spatial dimensions meaningfully map to the multi-dimensional model parameter space for empowering humans with new expressive tools? What are the limits of the spatial metaphor and in what ways (beyond proximity-dependent intent tags) should it interweave with or cede control to lexical or semantic modes of expression? What kind of collaboration becomes possible when both humans and AI co-create, reason, and reflect in a shared space? What new forms of hybrid visual & prompt literacy might emerge over time? We hope our work inspires future research on these and other questions, as well as further explorations of the *"intelligent use of space"* as shared, evolving sites of meaning—where both human and AI can leverage the rich implicit meaning of the canvas space itself as the prompt.

Acknowledgments

We thank the participants of our user study for taking the time to provide feedback on ImaginationVellum and the spatial prompting techniques, and the reviewers of this submission for concrete suggestions to improve this paper.

References

[1] Marah Abdin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J. Hewett, Mojan Javaheripi, Piero Kauffmann, James R. Lee, Yin Tat Lee, Yuanzhi Li, Weishung Liu, Caio C. T. Mendes, Anh Nguyen, Eric Price, Gustavo de Rosa, Olli Saarikivi, Adil Salim, Shital Shah, Xin Wang, Rachel Ward, Yue Wu, Dingli Yu, Cyril Zhang, and Yi Zhang. 2024. Phi-4 Technical Report. https://doi.org/10.48550/arXiv.2412.08905 arXiv:2412.08905 [cs].

[2] Affine. 2025. Affine app. https://affine.pro

[3] Stability AI. 2025. Stable Diffusion. https://github.com/Stability-AI

[4] 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 (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3613904.3642858

[5] Tyler Angert, Miroslav Ivan Suzara, Jenny Han, Christopher Lawrence Pondoc, and Hariharan Subramonyam. 2023. Spellburst: A Node-based Interface for Exploratory Creative Coding with Natural Language Prompts. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 1–22. https://doi.org/10.1145/3586183.3606719

[6] Ian Arawjo, Chelse Swoopes, Priyan Vaithilingam, Martin Wattenberg, and Elena Glassman. 2024. ChainForge: A Visual Toolkit for Prompt Engineering and LLM Hypothesis Testing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–18. https://doi.org/10.1145/3613904.3642016

[7] Seok-Hyung Bae, Ravin Balakrishnan, and Karan Singh. 2009. EverybodyLovesSketch: 3D sketching for a broader audience. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology (UIST '09)*. Association for Computing Machinery, New York, NY, USA, 59–68. https://doi.org/10.1145/1622176.1622189

[8] Mark Baskinger and William Bardel. 2013. *Drawing Ideas: A Hand-Drawn Approach for Better Design*. Watson-Guptill, New York, NY, USA.

[9] Olivier Bau and Wendy E. Mackay. 2008. OctoPocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st annual ACM symposium on User interface software and technology (UIST '08)*. Association for Computing Machinery, New York, NY, USA, 37–46. https://doi.org/10.1145/1449715.1449724

[10] D. Bauer, P. Fastrez, and J. Hollan. 2005. Spatial Tools for Managing Personal Information Collections. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. IEEE, USA, 104b–104b. https://doi.org/10.1109/HICSS.2005.551 ISSN: 1530-1605.

[11] Benjamin B. Bederson, James D. Hollan, Ken Perlin, Jonathan Meyer, David Bacon, and George Furnas. 1996. Pad++: A Zoomable Graphical Sketchpad For Exploring Alternate Interface Physics. *Journal of Visual Languages & Computing* 7, 1 (1996), 3–32. https://doi.org/10.1006/jvlc.1996.0002

[12] Margaret A. Boden. 2003. The creative mind: Myths and mechanisms: Second edition. *The Creative Mind: Myths and Mechanisms: Second Edition* (sep 2003), 1–344. https://doi.org/10.4324/9780203508527

[13] Aras Bozkurt. 2024. Tell Me Your Prompts and I Will Make Them True: The Alchemy of Prompt Engineering and Generative AI. *Open Praxis* 16, 2 (April 2024), 111–118. https://doi.org/10.55982/openpraxis.16.2.661

[14] 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

[15] Richard G. Brereton. 2009. *Sketchbooks: The Hidden Art of Designers, Illustrators and Creatives*. Laurence King Publishing, London, United Kingdom.

[16] Adam Brumm, Adhi Agus Oktaviana, Basran Burhan, Budianto Hakim, Rustan Lebe, Jian xin Zhao, Priyatno Hadi Sulistyarto, Marlon Ririmasse, Shinatria Adhityatama, Iwan Sumantri, and Maxime Aubert. 2021. Oldest cave art found in Sulawesi. *Science Advances* 7, 3 (2021), eabd4648. https://doi.org/10.1126/sciadv.abd4648

[17] Bill Buxton. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design* (1 ed.). Morgan Kaufmann, Burlington, Massachusetts, United States.

[18] Yining Cao, Yiyi Huang, Anh Truong, Hijung Valentina Shin, and Haijun Xia. 2025. Compositional Structures as Substrates for Human-AI Co-creation Environment: A Design Approach and A Case Study. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 188, 25 pages. https://doi.org/10.1145/3706598.3713401

[19] Stuart K. Card, George G. Robertson, and William York. 1996. The WebBook and the Web Forager: an information workspace for the World-Wide Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, British Columbia, Canada) *(CHI '96)*. Association for Computing Machinery, New York, NY, USA, 111–ff. https://doi.org/10.1145/238386.238446

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

[21] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3491102.3501819

[22] British Design Council. 2025. The Double Diamond. https://www.designcouncil.org.uk/our-resources/the-double-diamond/

[23] Ben Crothers. 2017. *Presto Sketching: The Magic of Simple Drawing for Brilliant Product Thinking and Design* (1st edition ed.). O'Reilly Media, Beijing Boston Farnham Sebastopol Tokyo.

[24] 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

[25] Hai Dang, Lukas Mecke, and Daniel Buschek. 2022. GANSlider: How Users Control Generative Models for Images using Multiple Sliders with and without Feedforward Information. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3491102.3502141

[26] Thomas Alva Edison. 2025. The Thomas Edison Papers. https://edison.rutgers.edu/

[27] Koos Eissen and Steur Roselien. 2019. *Sketching: Drawing Techniques for Product Designers* (reprint edition ed.). Laurence King Publishing, Amsterdam.

[28] Jennifer Fernquist, Tovi Grossman, and George Fitzmaurice. 2011. Sketch-sketch revolution: an engaging tutorial system for guided sketching and application learning. In *Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11)*. Association for Computing Machinery, New York, NY, USA, 373–382. https://doi.org/10.1145/2047196.2047245

[29] Marina Fernández Camporro and Nicolai Marquardt. 2020. Live Sketchnoting Across Platforms: Exploring the Potential and Limitations of Analogue and Digital Tools. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376192

[30] OpenJS Foundation. 2025. expressjs/express. https://github.com/expressjs/express

[31] OpenJS Foundation. 2025. node.js. https://nodejs.org/en

[32] Frederic Gmeiner, Nicolai Marquardt, Michael Bentley, Hugo Romat, Michel Pahud, David Brown, Asta Roseway, Nikolas Martelaro, Kenneth Holstein, Ken Hinckley, and Nathalie Riche. 2025. Intent Tagging: Exploring Micro-Prompting Interactions for Supporting Granular Human-GenAI Co-Creation Workflows. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 531, 31 pages. https://doi.org/10.1145/3706598.3713861

[33] Saul Greenberg, Sheelagh Carpendale, Nicolai Marquardt, and Bill Buxton. 2012. *Sketching User Experiences: The Workbook* (1 ed.). Morgan Kaufmann, Burlington, Massachusetts, United States.

[34] Joshua Hailpern, Erik Hinterbichler, Caryn Leppert, Damon Cook, and Brian P. Bailey. 2007. TEAM STORM: demonstrating an interaction model for working with multiple ideas during creative group work. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition* (Washington, DC, USA) *(C&C '07)*. Association for Computing Machinery, New York, NY, USA, 193–202. https://doi.org/10.1145/1254960.1254987

[35] Ken Hinckley, Shengdong Zhao, Raman Sarin, Patrick Baudisch, Edward Cutrell, Michael Shilman, and Desney Tan. 2007. InkSeine: *In Situ* search for active note taking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 251–260. https://doi.org/10.1145/1240624.1240666

[36] James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.* 7, 2 (June 2000), 174–196. https://doi.org/10.1145/353485.353487

[37] Stacy Hsueh, Marianela Ciolfi Felice, Sarah Fdili Alaoui, and Wendy E. Mackay. 2024. What Counts as 'Creative' Work? Articulating Four Epistemic Positions in Creativity-Oriented HCI Research. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 497, 15 pages. https://doi.org/10.1145/3613904.3642854

[38] Edwin Hutchins. 1996. *Cognition in the Wild*. Bradford Books; Revised ed. edition (September 1, 1996), St Bradford, PA, United States.

[39] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 17–24. https://doi.org/10.1145/642611.642616

[40] Takeo Igarashi, Satoshi Matsuoka, and Hidehiko Tanaka. 1999. Teddy: a sketching interface for 3D freeform design. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques (SIGGRAPH '99)*. ACM Press/Addison-Wesley Publishing Co., USA, 409–416. https://doi.org/10.1145/311535.311602

[41] Nanna Inie and Peter Dalsgaard. 2020. How Interaction Designers Use Tools to Manage Ideas. *ACM Trans. Comput.-Hum. Interact.* 27, 2 (March 2020), 7:1–7:26. https://doi.org/10.1145/3365104

[42] Peiling Jiang, Jude Rayan, Steven P. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3586183.3606737

[43] Gabe Johnson, Mark D. Gross, Jason Hong, and Ellen Yi-Luen Do. 2009. Computational Support for Sketching in Design: A Review. *Foundations and Trends® in Human–Computer Interaction* 2, 1 (Feb. 2009), 1–93. https://doi.org/10.1561/1100000013 Publisher: Now Publishers, Inc.

[44] Juriy Zaytsev, Stefan Kienzle, and Andrea Bogazzi. 2024. fabric.js. https://github.com/fabricjs/fabric.js

[45] Tae Soo Kim, Yoonjoo Lee, Minsuk Chang, and Juho Kim. 2023. Cells, Generators, and Lenses: Design Framework for Object-Oriented Interaction 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, 1–18. https://doi.org/10.1145/3586183.3606833

[46] David Kirsh. 1995. The intelligent use of space. *Artif. Intell.* 73, 1–2 (feb 1995), 31–68. https://doi.org/10.1016/0004-3702(94)00017-U

[47] Janin Koch, Nicolas Taffin, Michel Beaudouin-Lafon, Markku Laine, Andrés Lucero, and Wendy E. Mackay. 2020. ImageSense: An Intelligent Collaborative

Ideation Tool to Support Diverse Human-Computer Partnerships. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (May 2020), 45:1–45:27. https://doi.org/10.1145/3392850
[48] Janin Koch, Nicolas Taffin, Andrés Lucero, and Wendy E. Mackay. 2020. SemanticCollage: Enriching Digital Mood Board Design with Semantic Labels. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 407–418. https://doi.org/10.1145/3357236.3395494
[49] 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 arXiv:2303.12647.
[50] James A. Landay and Brad A. Myers. 1995. Interactive sketching for the early stages of user interface design. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95*. ACM Press, Denver, Colorado, United States, 43–50. https://doi.org/10.1145/223904.223910
[51] Tomas Lawton, Francisco J Ibarrola, Dan Ventura, and Kazjon Grace. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. ACM, Sydney NSW Australia, 264–277. https://doi.org/10.1145/3581641.3584095
[52] Yoonjoo Lee, John Joon Young Chung, Tae Soo Kim, Jean Y Song, and Juho Kim. 2022. Promptiverse: Scalable Generation of Scaffolding Prompts Through Human-AI Hybrid Knowledge Graph Annotation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3502087
[53] Antonios Liapis, Georgios N. Yannakakis, Constantine Alexopoulos, and Phil Lopes. 2016. Can Computers Foster Human Users' Creativity? Theory and Praxis of Mixed-Initiative Co-Creativity. *Digital Culture and Education* 8, 2 (jul 2016). https://doaj.org/article/5674a85e47fb452fa6cbfe279aa271ba
[54] 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, Article 427, 18 pages. https://doi.org/10.1145/3706598.3713397
[55] Siân Lindley, Xiang Cao, John Helmes, Richard Morris, and Sam Meek. 2013. Towards a Tool for Design Ideation: Insights From Use of SketchStorm. In *BCS-HCI '13: Proceedings of the 27th International BCS Human Computer Interaction Conference*. BCS-HCI, United Kingdom. https://doi.org/10.5555/2578048.2578068
[56] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA, Article 1516, 25 pages.
[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 (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3544549.3577043
[58] 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 (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3613904.3642861
[59] J. Marks, B. Andalman, P. A. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. Shieber. 1997. Design galleries: a general approach to setting parameters for computer graphics and animation. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH '97)*. ACM Press/Addison-Wesley Publishing Co., USA, 389–400. https://doi.org/10.1145/258734.258887
[60] 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 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
[61] Justin Matejka, Michael Glueck, Erin Bradner, Ali Hashemi, Tovi Grossman, and George Fitzmaurice. 2018. Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173943
[62] Scott McCloud. 1994. *Understanding Comics: The Invisible Art* (reprint edition ed.). William Morrow Paperbacks, New York.
[63] Midjourney. 2025. Midjourney. https://www.midjourney.com/website
[64] Bryan Min and Haijun Xia. 2025. Feedforward in Generative AI: Opportunities for a Design Space. https://doi.org/10.48550/arXiv.2502.14229
[65] Miro. 2025. Miro app. https://miro.com
[66] Aditi Mishra, Bretho Danzy, Utkarsh Soni, Anjana Arunkumar, Jinbin Huang, Bum Chul Kwon, and Chris Bryan. 2025. PromptAid: Visual Prompt Exploration, Perturbation, Testing and Iteration for Large Language Models. *IEEE Transactions on Visualization and Computer Graphics* (2025), 1–14. https://doi.org/10.1109/TVCG.2025.3535332
[67] Arefin Mohiuddin and Robert Woodbury. 2020. Interactive Parallel Coordinates for Parametric Design Space Exploration. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3383101
[68] Thomas P. Moran, Patrick Chiu, and William van Melle. 1997. Pen-based interaction techniques for organizing material on an electronic whiteboard. In *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology* (Banff, Alberta, Canada) *(UIST '97)*. Association for Computing Machinery, New York, NY, USA, 45–54. https://doi.org/10.1145/263407.263508
[69] Elizabeth D. Mynatt, Takeo Igarashi, W. Keith Edwards, and Anthony LaMarca. 1999. Flatland: new dimensions in office whiteboards. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 346–353. https://doi.org/10.1145/302979.303108
[70] Kumiyo Nakakoji, Yasuhiro Yamamoto, Shingo Takada, and Brent N. Reeves. 2000. Two-dimensional spatial positioning as a means for reflection in design. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (New York City, New York, USA) *(DIS '00)*. Association for Computing Machinery, New York, NY, USA, 145–154. https://doi.org/10.1145/347642.347697
[71] Obsidian. 2025. Obsidian Canvas app. https://obsidian.md/canvas
[72] Timothy O'Donnell. 2009. *Sketchbook: Conceptual Drawings from the World's Most Influential Designers*. Rockport Publishers, Beverly, MA, USA.
[73] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I Lead, You Help but Only with Enough Details: Understanding User Experience of Co-Creation with Artificial Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174223
[74] Daniel Orban, Daniel F. Keefe, Ayan Biswas, James Ahrens, and David Rogers. 2019. Drag and Track: A Direct Manipulation Interface for Contextualizing Data Instances within a Continuous Parameter Space. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 256–266. https://doi.org/10.1109/TVCG.2018.2865051 Conference Name: IEEE Transactions on Visualization and Computer Graphics.
[75] Srishti Palani and Gonzalo Ramos. 2024. Evolving Roles and Workflows of Creative Practitioners in the Age of Generative AI. In *Proceedings of the 16th Conference on Creativity & Cognition (C&C '24)*. Association for Computing Machinery, New York, NY, USA, 170–184. https://doi.org/10.1145/3635636.3656190
[76] Elin Rønby Pedersen, Kim McCall, Thomas P. Moran, and Frank G. Halasz. 1993. Tivoli: an electronic whiteboard for informal workgroup meetings. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 391–398. https://doi.org/10.1145/169059.169309
[77] Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2024. DesignPrompt: Using Multimodal Interaction for Design Exploration with Generative AI. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 804–818. https://doi.org/10.1145/3643834.3661588
[78] 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, Article 874, 20 pages. https://doi.org/10.1145/3706598.3714027
[79] Morgan N. Price, Gene Golovchinsky, and Bill N. Schilit. 1998. Linking by inking: trailblazing in a paper-like hypertext. In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia: Links, Objects, Time and Space—Structure in Hypermedia Systems: Links, Objects, Time and Space—Structure in Hypermedia Systems* (Pittsburgh, Pennsylvania, USA) *(HYPERTEXT '98)*. Association for Computing Machinery, New York, NY, USA, 30–39. https://doi.org/10.1145/276627.276631
[80] Gonzalo Ramos, Rick Barraza, Victor Dibia, and Sharon Lo. 2024. The Aleph & Other Metaphors for Image Generation. *ArXiv* abs/2402.07104 (2024). https://arxiv.org/abs/2402.07104
[81] Patrick Renner, Thies Pfeiffer, and Ipke Wachsmuth. 2014. Spatial References with Gaze and Pointing in Shared Space of Humans and Robots. In *Spatial Cognition IX*, Christian Freksa, Bernhard Nebel, Mary Hegarty, and Thomas Barkowsky (Eds.). Springer International Publishing, Cham, 121–136. https://doi.org/10.1007/978-3-319-11215-2_9
[82] Nathalie Riche, Anna Offenwanger, Frederic Gmeiner, David Brown, Hugo Romat, Michel Pahud, Nicolai Marquardt, Kori Inkpen, and Ken Hinckley. 2025. AI-Instruments: Embodying Prompts as Instruments to Abstract & Reflect Graphical Interface Commands as General-Purpose Tools. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1104, 18 pages.

https://doi.org/10.1145/3706598.3714259
[83] Yann Riche, Nathalie Henry Riche, Ken Hinckley, Sheri Panabaker, Sarah Fuelling, and Sarah Williams. 2017. As We May Ink? Learning from Everyday Analog Pen Use to Improve Digital Ink Experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3241–3253. https://doi.org/10.1145/3025453.3025716
[84] Dan Roam. 2013. *The Back of the Napkin (Expanded Edition): Solving Problems and Selling Ideas with Pictures* (expanded edition ed.). Portfolio, New York, NY, USA.
[85] Hugo Romat, Nathalie Henry Riche, Ken Hinckley, Bongshin Lee, Caroline Appert, Emmanuel Pietriga, and Christopher Collins. 2019. ActiveInk: (Th)Inking with Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300272
[86] Bill N. Schilit, Gene Golovchinsky, and Morgan N. Price. 1998. Beyond paper: supporting active reading with free form digital ink annotations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI '98)*. ACM Press/Addison-Wesley Publishing Co., USA, 249–256. https://doi.org/10.1145/274644.274680
[87] Donald A. Schon and Glenn Wiggins. 1992. Kinds of seeing and their functions in designing. *Design Studies* 13, 2 (April 1992), 135–156. https://doi.org/10.1016/0142-694X(92)90268-F
[88] D. A. Schön. 1992. Designing as reflective conversation with the materials of a design situation. *Knowledge-Based Systems* 5, 1 (March 1992), 3–14. https://doi.org/10.1016/0950-7051(92)90020-G
[89] Lei Shi, Rachel Campbell, Peggy Chi, Maria Cirimele, Mike Cleron, Kirsten Climer, Chelsey Q Fleming, Ashwin Ganti, Philippe Gervais, Pedro Gonnet, Tayeb A Karim, Andrii Maksai, Chris Melancon, Rob Mickle, Claudiu Musat, Palash Nandy, Xiaoyu Iris Qu, David Robishaw, Angad Singh, and Mathangi Venkatesan. 2024. Inkeraction: An Interaction Modality Powered by Ink Recognition and Synthesis. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–26. https://doi.org/10.1145/3613904.3642498
[90] Ben Shneiderman. 1999. User interfaces for creativity support tools. In *Proceedings of the 3rd conference on Creativity & cognition*. ACM, New York, NY, USA, 15–22.
[91] 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
[92] Pao Siangliulue, Joel Chan, Steven P Dow, and Krzysztof Z Gajos. 2016. IdeaHound: improving large-scale collaborative ideation with crowd-powered real-time semantic modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 609–624.
[93] Beat Signer. 2025. Pen-Based Interaction. In *Handbook of Human Computer Interaction*, Jean Vanderdonckt, Philippe Palanque, and Marco Winckler (Eds.). Springer Nature Switzerland, Cham, 1–22. https://doi.org/10.1007/978-3-319-27648-9_102-1
[94] StabilityAI. 2024. CompVis/stable-diffusion · Hugging Face. https://huggingface.co/CompVis/stable-diffusion
[95] Hari Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs. 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/10.1145/3613904.3642754
[96] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–18. https://doi.org/10.1145/3586183.3606756
[97] Ivan E. Sutherland. 1964. Sketch pad a man-machine graphical communication system. In *Proceedings of the SHARE design automation workshop (DAC '64)*. Association for Computing Machinery, New York, NY, USA, 6.329–6.346. https://doi.org/10.1145/800265.810742
[98] Michael Terry and Elizabeth D. Mynatt. 2002. Recognizing creative needs in user interface design. In *Proceedings of the 4th Conference on Creativity & Cognition* (Loughborough, UK) *(C&C '02)*. Association for Computing Machinery, New York, NY, USA, 38–44. https://doi.org/10.1145/581710.581718
[99] Barbara Tversky. 2002. What do Sketches Say about Thinking? *AAAI Technical Report SS-02-08* (2002), 4.
[100] Jo Vermeulen, Kris Luyten, Elise van den Hoven, and Karin Coninx. 2013. Crossing the bridge over Norman's Gulf of Execution: revealing feedforward's true identity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1931–1940. https://doi.org/10.1145/2470654.2466255
[101] Samangi Wadinambiarachchi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3613904.3642919
[102] Graham Wallas. 1926. *The art of thought.* Harcourt, Brace, USA.
[103] 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 (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 136–146. https://doi.org/10.1145/3563657.3596072
[104] 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 (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3613904.3642803
[105] Jiafu Wei, Chia-Ming Chang, Xi Yang, and Takeo Igarashi. 2024. CanvasPic: An Interactive Tool for Freely Generating Facial Images Based on Spatial Layout. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3613905.3650952
[106] Pierre Wellner. 1993. Interacting with paper on the DigitalDesk. *Commun. ACM* 36, 7 (July 1993), 87–96. https://doi.org/10.1145/159544.159630
[107] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–22. https://doi.org/10.1145/3491102.3517582
[108] Haijun Xia, Ken Hinckley, Michel Pahud, Xiao Tu, and Bill Buxton. 2017. WritLarge: Ink Unleashed by Unified Scope, Action, & Zoom. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3227–3240. https://doi.org/10.1145/3025453.3025664
[109] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3544548.3581388
[110] Robert Zeleznik, Andrew Bragdon, Ferdi Adeputra, and Hsu-Sheng Ko. 2010. Hands-on math: a page-based multi-touch and pen desktop for technical work and problem solving. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 17–26. https://doi.org/10.1145/1866029.1866035
[111] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Paris, France, 3813–3824. https://doi.org/10.1109/ICCV51070.2023.00355