Protosampling: Enabling Free-Form Convergence of Sampling and Prototyping through Canvas-Driven Visual AI Generation

Alicia Guo
Autodesk Research
Toronto, Ontario, Canada
University of Washington
Seattle, Washington, USA
axguo@uw.edu

David Ledo
Autodesk Research
Toronto, Ontario, Canada
david.ledo@autodesk.com

George Fitzmaurice
Autodesk Research
Toronto, Ontario, Canada
george.fitzmaurice@autodesk.com

Fraser Anderson
Autodesk Research
Toronto, Ontario, Canada
fraser.anderson@autodesk.com

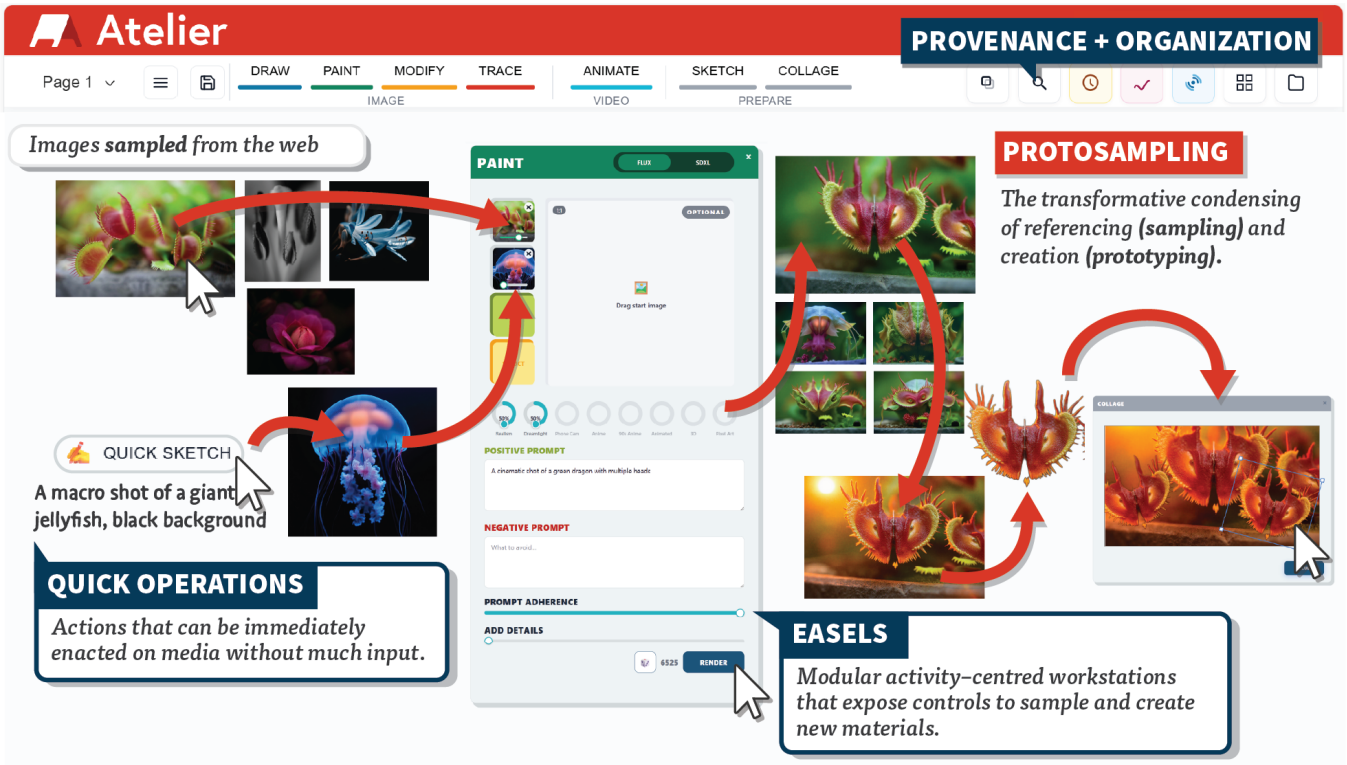

Figure 1: Atelier is a system that operationalizes protosampling - the practice of tightening sampling and prototyping for creating AI visuals. It enables a co-existence between collected and generated assets, provides a set of 'easels' that encapsulate complex workflows, and fosters collection and reflection through provenance.

Abstract

As an emergent process, creativity relies on explorations via sampling and prototyping for problem construction. These activities compile knowledge, provide a context enveloping the solution, and answer questions. With Generative AI, practitioners can go beyond sampling existing media towards instantly generating and remixing new ones. We refer to this convergence as 'protosampling'. Using existing literature we ground a definition for protosampling and operationalize it through Atelier, a canvas-like system that leverages a variety of generative image and video models for visual creation. Atelier: (1) blends the spaces for thinking and creation, where both references and generated assets co-exist in one space, (2) provides

This work is licensed under a Creative Commons Attribution 4.0 International License.

CHI '26, Barcelona, Spain

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791884

various encapsulated technical workflows that focus on the activity at hand, and (3) enables navigating emergence through interactive visualizations, smart search, and collections. Protosampling as a lens reframes creative work to emphasize the process itself and how seemingly disjointed thoughts can tightly interweave into a final solution.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**.

Keywords

Creativity Support Tools, Generative AI and Creativity, AI images and Videos, Sampling, Prototyping, Canvas-Based Interactions

ACM Reference Format:
Alicia Guo, David Ledo, George Fitzmaurice, and Fraser Anderson. 2026. Protosampling: Enabling Free-Form Convergence of Sampling and Prototyping through Canvas-Driven Visual AI Generation. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 27 pages. https://doi.org/10.1145/3772318.3791884

1 Introduction

The accelerated growth of generative AI enables new capabilities for generating rich media (e.g., images, videos, 3D models, audio) primarily using natural language . Over time, novel interactions enhance expressiveness and control by (1) providing multi-modal inputs, (e.g., using images to guide a new generated image's structure [122]), and (2) exposing relevant parameters that can affect the resulting generation. Generative AI interfaces exist on a broad spectrum catering to various audiences, where on one end simplified chat-like systems leverage conversational language to provide quick solutions. Some tools expose multiple degrees of inputs and controls such as MidJourney[71], where one can type a prompt, select the aspect ratio, provide a style reference, etc. At the other end of the spectrum, powerful programmatic tools offer the full gamut of models, inputs, and parameter settings, requiring specialized expertise. One popular system is ComfyUI [25], which offers a node-based programming approach to simultaneously combine and control multiple models. These systems encapsulate key atomic operations into nodes that can connect multiple models, algorithms, and even online services. While rich and expressive, this paradigm focuses primarily on procedural thinking for problem solving, and requires sourcing and selecting among many models.

This spectrum leaves a considerable gap for creative practitioners who want a balance of control while working in a medium that enables them to best express their intent. This means bringing their media and their process center stage, and making tools accessible when and where they are needed[17]. By focusing on the creative process, it becomes possible to harness a latent quality of generative tools – the ability to combine a wide variety of thinking materials and to use them to generate *new materials*, continuing engagement in a vast emergent recombination [26], a true conversation with the creative problem at hand [95]. This posits an interweaving of two key creative activities: sampling and prototyping. Viewing these two activities in tandem enables rethinking the creative medium as one where practitioners use the creative problem as a lens to look at the world, curate information, and actively transform it into partial solutions that inform the final result over time. To explore the interplay between sampling and prototyping, we designed Atelier (Figure 1, a canvas-like interface that brings together source material and generated content in the same space, and a set of activity-centric widgets offering generative AI operations with carefully selected inputs and parameters.

Atelier provides mechanisms to engage with content via collections, search, history, and interactive visualizations that enable reflection and reinterpretation of the body of work to surface emergent traits within the process. An extended 4-hour first-use study with 5 creative professionals further demonstrates Atelier's expressiveness and how it might be used in real-world scenarios. Specifically, this paper proposes the following contributions:

(1) A characterization of 'Protosampling' – a lens that treats sampling and prototyping as a joint activity rather than separate stages of the creative process.
(2) Atelier, a canvas-like system that operationalizes Protosampling to support the generation of media, with a particular focus on visuals (images and video).
(3) A set of abstractions and novel workflows for generating AI images and videos while focusing on creative controllability, demonstrated via a set of usage scenarios.

2 Protosampling

The creative process is often described as a discrete set of *stages* practitioners undergo when solving a creative problem [49, 94], with myriads of representations going from problem to solution [1, 27, 49, 54, 94]. For example, Sawyer's 8-stage model consists of: ask (find the problem), learn (acquire the knowledge), look (gather relevant information), play (incubation), think (generate ideas), fuse (combine ideas), choose (select ideas) and make (externalize ideas) [94]. Sawyer's model of creativity amalgamates and extends many past creative process models including popular ones in design such as IDEO [16, 18, 40, 46, 50, 88, 96, 103, 108]. Early stages emphasize thinking and conceptualization which are typically associated with sampling, while later stages emphasize idea selection and creation, typically associated with prototyping.

One challenge is that these stage-based models attempt to systematize a creative process that is inherently messy and nonlinear, which has temporal overlaps and non-deterministic ordering [94]. In practice, these non-systemic activities are done in parallel where both problem and solution are actively reformulated[27]. Thus, Sawyer argues that the stages can be better interpreted as disciplines or "habits of mind" followed by practitioners throughout their work. Creative activities are a result of "mini-insights" [94] that take place throughout the whole process [9]; they are often interrupted, reconstructed, and repeated [89], with evidence demonstrating non-linearity [92, 106].

An alternative way to look at the creative process is to consider different abstractions that are not captured in these types of representations. For instance, literature considers activities such as problem construction, sampling, and prototyping. Specifically, we look at two activities not often discussed in conjunction: ***sampling*** as a way of accessing insights through *knowledge* and ***prototyping***

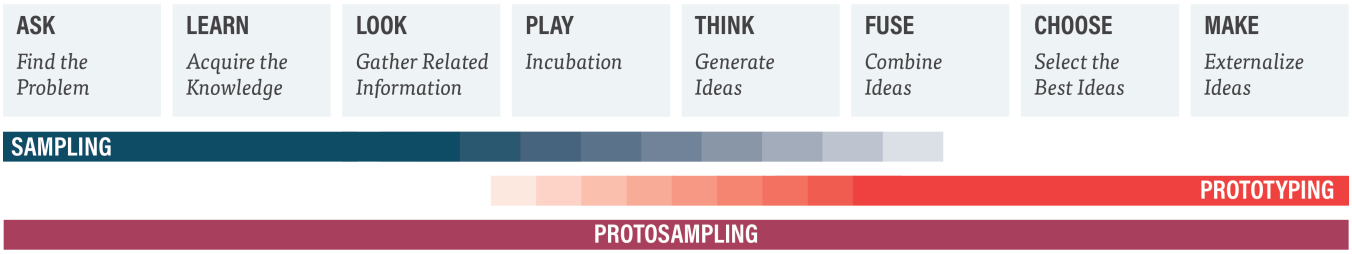

Figure 2: Overview of the creative process according to Sawyer [94], situating sampling, prototyping and proposing Protosampling as the bridge between these two activities.

as a way of accessing insights through *making* (Figure 2). While prototyping is sometimes listed as an explicit step withing the process, these activities encompass more than one stage/discipline. For example, both sampling and prototyping can help with generating ideas, but do so in different ways. Sampling and prototyping are activities that cut across the stages mentioned in previous creativity models and provide a new lens from which to look at the life cycle of ideas and material creations.

Sampling refers to *the act of collecting, organizing and transforming materials* [101]. Creative practitioners engage in sampling the real world [41, 64], continuously collecting materials and information, ranging from inspiration, relevant bits, to the problem at hand. Sampling is a form of *'opportunistic assimilation'*: people going about their day encounter objects or situations relevant to an unsolved problem [61]. The creative problem then becomes a lens from which to look at the world, creators look at information in their environment to link new information and integrate it into existing problems and tasks. For example, a musician might capture a soundscape from an environment, trim and edit the audio to use as a beat in a piece, or deconstruct a song's chord progression which inspires further composition. A graphic designer might keep a postcard because they like the colours. Ultimately, materials are collected, curated, in some cases formalized (e.g., mood boards) and then used. As part of creative problem-solving, Stemasov et al. [101] characterize sampling from past literature as an act that structures thinking while being open-ended and deliberate. Sampling provides context around creative work, synthesizes collections into new designs, offers potential raw materials, collects design decisions, and helps reinterpret the problem.

Prototyping, on the other hand, focuses on actual creation to better understand a problem. Prototypes are seen as *purposeful manifestations of design ideas that traverse a design space, leading to meaningful knowledge about some aspects of a final design* [59]. Prototyping can serve different roles, such as exploration, experimentation, or evolution [34]. Of particular interest is the ability to explore design spaces – a "fuzzy" boundary between exploration and experimentation [34]. While Buxton distinguishes *sketches* from *prototypes* [19], much of the design literature [27, 39, 59, 62] advocates for prototyping the same way Buxton frames sketching: prototypes enable exploring design spaces, testing out alternatives, and informing rationales [59]. Stolterman argues that sketching is a rigorous approach to explore and iterate ideas. Prototypes are strong *because* they are incomplete solutions, acting as *'filters'* that examine the quality of an idea and answer a specific question without having to create the final design [59]. Existing work often frames prototyping in terms of exploring specific areas of a design such as function, structure, behaviour, look and feel, role, implementation, etc. [26, 38, 59]. Lim et al. articulate how prototypes examine properties of the design space [59], for example exploring appearance by trying colours or shapes. While the term 'prototyping' is often used in design contexts, it applies to creative practice more broadly. Prototyping in animation might resemble different activities: *storyboards* examine camera angles and how the story translates to visuals, and *animatics* explore timing. To create the storyboard, an artist might make small *thumbnails* to visualize composition. This is why Logan and Smithers [62] argue that prototypes are all interconnected.

In a broad sense, both sampling and prototyping are exploratory activities enabling practitioners to engage in ***"problem construction"*** [49], a process in which open-ended elements of the problem are tightened and reframed to drive the solution. Cross [27] describes problem construction as a process beginning with ambitious high level goals, followed by periods of intense activity leading to reflective contemplation compounds new understanding to redefine the problem: the problem and solution co-evolve.

The Impact of Generative AI. While there is still limited understanding of generative AI's implications for the creative process, the reality is that it has fundamentally changed it. Media such as text, images, videos, 3D models, music, etc. can be instantiated rapidly. Materials can be remixed and reused in new generations. Furthermore, this process means many more partial solutions get created in shorter bursts of time. For example, Ledo's work [56] shows how generating a seconds-long animation required over 600 generations. This type of AI-driven creation is bound to encompass snapshots of the process, with each generation embodying parts of the understanding of the problem at that given point. *Sampling has become more like prototyping*, as assets are not only collected but also created, *and prototyping is becoming more like sampling*, as both collected and generated materials can be combined and reused. Seeing this loop becoming tighter, we see value in bridging these two concepts together: *Protosampling*.

2.1 Procedural Action

To enact Protosampling, it is necessary to operationalize the connection between the *sense-making* from collected information to the *action* of externalizing. Fortunately, past theories and studies have abstracted many of these principles. Mental representations

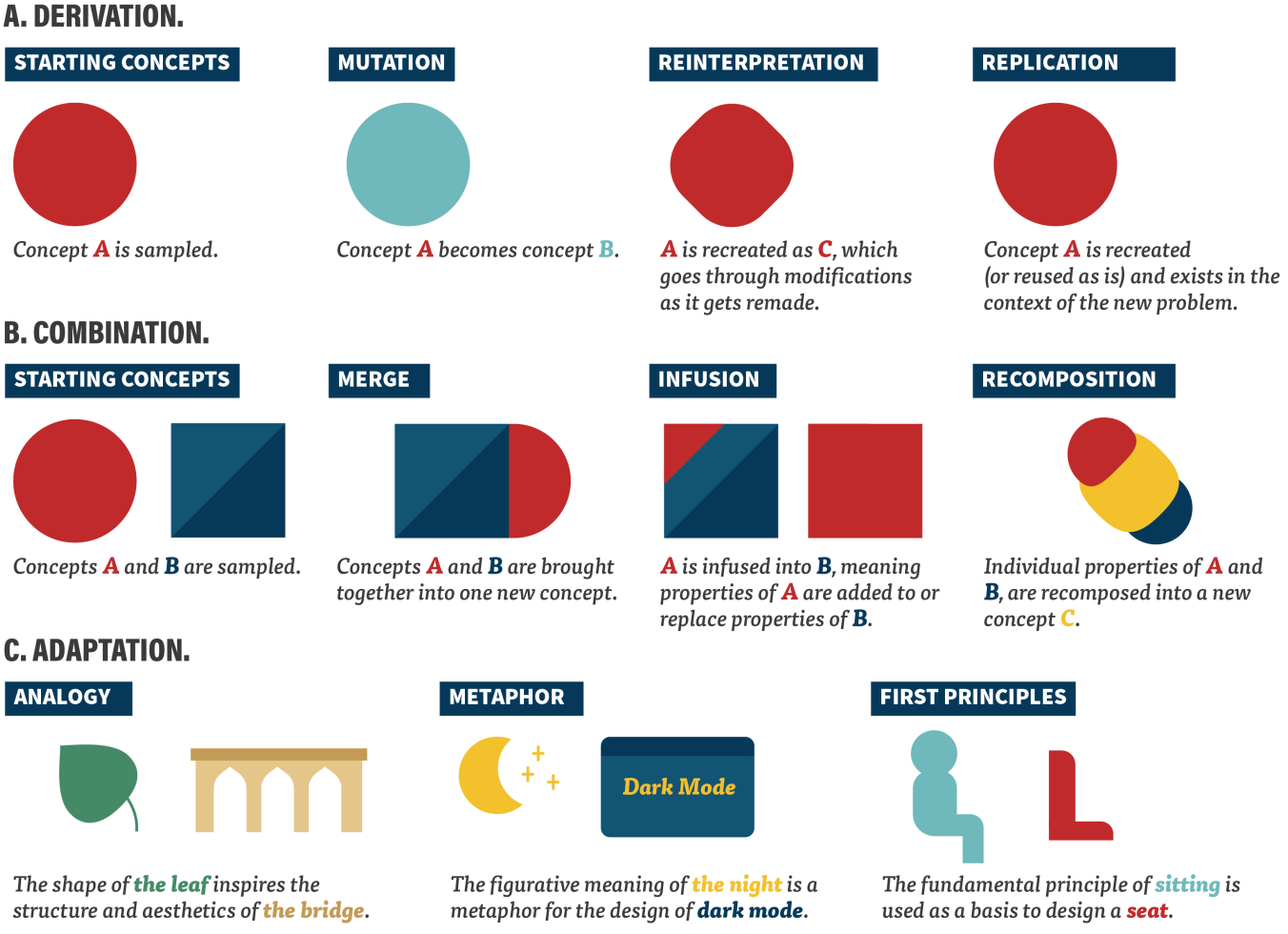

Figure 3: Procedural actions within Protosampling, these include (a) derivations, in which concept properties are changed; (b) combinations, in which two concepts are brought together; and (c) adaptation, in which concepts are transferred from other guiding principles.

are referred to as concepts, and they have properties with associated values [12, 99] (e.g., a ball has a size, colour, etc.). We outline and categorize different ways to manifest ideas, which provides a lens to understand Protosampling (Figure 3).

2.1.1 Derivation. Gero [38] as well as Cross[26] outline incremental changes an idea can undergo. We also account for how ideas might integrate into a creative problem [66, 102]. It is worth noting that these categories for procedural actions in Protosampling are not mutually exclusive.

Mutation. An incremental change that transforms a concept. This can involve one or more properties that are altered. For example, an object is brought in and its colours are changed.

Reinterpretation. A recreation of a concept, which in the process of recreation gets modified. Specific properties may change in this process. One example is a musical piece that gets reharmonized or played with new instruments.

Replication. Either a faithful recreation or reuse of an existing asset. Most properties are likely intact. For instance, assets from one project are brought in and reused in a new production.

2.1.2 Combination. One way in which ideas progress in creative processes is combination. The literature uses *conceptual combination* as "a mental act by which imagination brings concepts together to produce new ideas in creative processes" [45]. Creative theories often focus on language, and have derived six types of combination [94]. For simplicity and to make combination more actionable, we distill 3 types of combination to expand on existing takes [26, 38]:

Merge. Two sampled concepts are brought together into the context of the creative problem. The two concepts co-exist together.

For example, images are brought together in preparation for a newspaper article.

Infusion. Given two concepts, the properties of one concept are transferred to another concept, either changing its properties and/or adding new ones. For example, the colour palette of an image is used to set the colour for the typography on a page.

Recomposition. Two concepts are taken apart and the properties are used in the creation of a new concept. For instance, different samples in music can be extracted from two different songs and reused in a new composition. Note how recomposition implies a deconstruction of the source material for reassembly.

2.1.3 Adaptation. Creative work often benefits from bringing concepts from one domain into another [94]. What distinguishes adaptation from derivation is the source and how it is used within the context of the creative problem.

Analogy. It is a structured comparison transferring functions, logic or relationships into a domain: the identifying of patterns/threads across ideas [23]. An example is a subway map, which translates physical locations to only show abstracted relevant information.

Metaphor. It is a figurative framing that transfers meaning from one domain to another [94]. Unlike a combination, a metaphor relies on a concept that is not directly linked to the topic at hand [85]. An example is skeuomorphism in HCI/UX design, such as folders and the file system.

First Principles. First principles are basic truths and assumptions about a situation that cannot be reduced further, a return to the fundamentals [26]. For example, a first principle when thinking of a chair is that it supports a body at rest.

2.2 Situatedness

Protosampling can be considered a situated activity that takes place in a workspace (whether physical or virtual). Traditionally, sampling is typically associated with mood boarding [64]. While creativity theory emphasizes the ideation and thinking part of creativity, and design theory describes prototyping as a practice, these domains rarely discuss what happens with the work itself. Fortunately, we can draw from studies on knowledge work to further understanding of the creative process, as creative practitioners are a significant subset of knowledge workers[13].

Knowledge workers process information to generate new knowledge [51]. Like the creative process, knowledge work is described as erratic and non-deterministic [14], with individuals working with many materials at a given time. Activities can range from highly unstructured tasks that emphasize information to more structured ones that emphasize documentation [13]. The key knowledge work tasks tie to the creative process, as an interplay of capturing, organizing, formalizing, and retrieving [83]. Materials are constantly being used and remixed [42].

Knowledge work studies also emphasize engagement with the workspace – depending on the task, information is organized (e.g., into folders) or grouped in piles [13, 67], with the workspace actively providing context, where layout acts as a living reminder of the current activity. Materials are distinguished by appearance, size and position, proximity to tools, and overall placement. These implicitly provide clues about the current task. For example, a printed document can be annotated and placed on a keyboard with a pen on top. When the practitioner walks away, this document acts as a *trail* of information on the current state: one can see what the document is, the number of pages, the progress on annotations, and the placement on the keyboard indicates it is active. The manipulability of the physical document (flipping through pages) brings context at no cost [14], and the need to apply mechanisms such as search decrease when a document is *'hot'* [97]. Logan and Smithers argue that materials and prototypes serve to evoke memory, leading to progress towards a solution. Meanwhile, less relevant information gets stored to reduce clutter [42, 97], reducing its visibility – it becomes searchable yet easily lost [42], retrieved only for specific needs [48] in a given context [84]. For creative practitioners, how information is managed highlights the value of structuring their search and keeping relevant information handy[74].

It is no surprise then that creative practitioners have adopted workarounds to emulate physical spaces when working with digital tools. Frich et al. [36] describe the use of margins around digital workspaces to retain relevant materials and past versions. This suggests an active use of Protosampling across domains, and helps explain a recurring desire to create tools for accessing and manipulating information *"at the right place and the right time"* [17].

This perspective highlights the challenge posed by fragmented information in digital tools. Following the ***trails*** of a process holds additional value often missed with the active digitization of materials. Such concern is especially relevant to *incubation*, as insights are often derived when stepping away from the work [27, 94].

3 Related Work

This work builds upon prior systems that support sampling for ideation and generative AI interfaces that provide a variety of interactions between inputs and outputs.

3.1 Materials as Inspiration and Samples

Digital and physical mood boards allow creators to explore design spaces, visualize ideas, and share their visions with others through collecting reference material arranged in a spatial manner [52]. These serve to structure the problem and align on aesthetic or conceptual directions [65], and become more effective when treated as active artifacts that evolve with the project [70]. Recent work frames sampling as a core creative activity where creators capture and collect materials, and even organize and remix them. Moodcubes [47] and VRicolage [101] are examples going beyond flat mood boards, leveraging spatial and multi-modal sampling to encourage recombinating references.

Generative AI allows for rapid generation of inspiration material [87, 110] and supports quick ideation in many visual domains [7, 81, 112]. GanCollage [110] integrates generative AI into mood boards, tagging generated images with semantic labels used to explore the latent visual space. These interactions focus on exploring variations rather than on composing or editing artifacts. Other systems augment mood boards with semantic tagging and search to

navigate and reflect on spaces with less visual organizational structure [52, 53, 87]. These works primarily treat found and generated artifacts as references for later use in the creative process.

3.2 Creating Media with Generative AI Interfaces

Generative AI systems vary in how they treat inputs and outputs, the generation process, and interaction modalities.

Prompt-Based Interfaces. Systems such as ChatGPT [82] and Midjourney [71] primarily use natural language prompting as an accessible method to generate AI outputs. These interfaces present a linear thread of prompts and results that support a single active line of exploration. Outputs are chronologically arranged, burying previous generations in history. While approachable, these interfaces lack control, discoverability of operations, and navigation.

Node-Based Interfaces. On the other end of the spectrum systems like ComfyUI [25] expose models' internal parameters through a node-based interface and allow chaining modules together into workflows for generating media. These offer higher degrees of control, through still focus on executing one workflow at a time. Final outputs can then be piped back as inputs for a new workflow, where the intermediate steps are erased. Node-based interfaces prioritize visible components for *atomic operations*, where nodes can load files, run operations, or save outputs.

Some systems look to make node-based programming more approachable, such as FLORA [33], Runway workflows [90], and Figma Weave [32]. These tools turn media into nodes, enabling use of multiple images and revealing outputs as they are processed within an explicit graph structure. The *links* between objects are emphasized, highlighting the *operations* taking place. The thinking processes still primarily focuses on graph-based workflows, where outputs are cleared and re-generated every time the workflow is executed. FLORA describes its own approach as *"built for professionals who think in systems, not just outputs"* [33].

Bringing Direct Manipulation to AI Generation. Between these two extremes of emphasizing prompts versus operations, some systems provide direct manipulation and extend prompts with other inputs, such as sketching [24, 60, 91, 120], image references [58, 104], region masking [28], or composable prompt widgets [6].

3.3 Virtual Canvases

Canvas interfaces are 2D planes where content can be imported, created, or modified, including those for brainstorming and collaboration (e.g., Mural[76], Miro[73]), and others optimized for note-taking (e.g., MilaNote [72]). Canvases enable arranging media in space, maintaining visibility and supporting branching, nonlinear exploration[52]. Other work explores attributes (such as shape or color) as first-class objects that can be manipulated[117], supporting free-form drawing and collecting resources [43] or coupling media with tools[17, 109, 117].

With generative AI, new canvas interfaces have started to emerge. ImaginationVellum [68] uses 'generative strokes' to rapidly create design variations by leveraging spatial relationships to text to modulate between concepts. Adobe Firefly Boards[3] feature canvases for organizing and displaying generated content, providing a global prompt-based textbox with additional controls.

3.4 Towards Operationalizing Protosampling

Generative AI has begun to reshape creative workflows across visual domains[22, 63, 98, 121], accelerating exploration and rapid generation of artifacts. The speed and capabilities of generative AI have evolved to condense sampling and prototyping, merging thinking and making into a more closely coupled process. Systems such as Firefly Boards[3], Paratrouper[58] and ImaginationVellum[68] already lean towards that convergence. Firefly Boards provides various online models, and its image generation can use one style image and one structure reference. Paratrouper [58] allows for combination and derivation while providing more structured exploration and hiding away reference material. ImaginationVellum[68] excels at derivation and blending sketches into images. However, these emphasize particular aspects of the process rather than integrating sampling and prototyping as a unified spatial practice.

With Atelier, we operationalize Protosampling through a media-first canvas providing access to a variety of local open-source models and algorithms to manipulate and generate content. Interaction is presented in modular widgets abstracting activities (Draw, Paint, Trace, Modify, Animate) so creators can focus on the task rather than model selection or workflow authoring. At the same time, these widgets reveal every available input that can be used, and materials – both inputs and outputs co-exist in the same visual space. Thus, we restrict the high flexibility of a node-based paradigm while making a wide variety of parameters to be accessed and tweaked. We also translate terms to be actionable for creators without being overly technical, and hide features that require more expertise, such as sampler and a scheduler selection. Atelier's novelty lies in these interactions and the abstractions.

4 Design Rationale

To develop a system enacting Protosampling for visual media, we synthesized (1) learnings from prior work on canvas-based creative tools and generative AI workflows (§2-3), (2) continuous testing with newly released models and tools, (3) sustained participation in online creator communities [2, 10] and viewing of publicly shared video processes and tutorials[75, 107], and (4) our experiences testing the system as we prototyped and iterated on it[57, 78, 100, 124].

From these, we iteratively identified challenges and common creative strategies, such as creators moving between a host of tools in their process, and used these observed patterns to formulate our design rationale for Protosampling.

R1. Blending Spaces for Thinking and Making. Tightening the loop between sampling and prototyping requires a workspace that invites open-ended exploration and creation while also bringing materials together. Materials should be able to be brought in, generated, and recombined. This motivates a media-first interface that takes advantage of spatial arrangements for meaning and organization, leading to a canvas-like interaction. This also supports hot and cold areas [97], where creators can pile information according to their needs. To invite equal usage of sampled and generated media, they should be indistinguishable.

R2. Encapsulating Technical Workflows by Activity while Offering Control. Working with multiple types of generative AI models typically requires creators to determine which model to use along with settings and parameters. This process can be quite technical and break the creators flow. To enable smooth combination of multiple media types quickly, certain aspects of the model should be abstracted away. It is important to determine which parameters to expose so that creators have agency and control, but need not spend significant time learning or tweaking parameters. The interface must expose the right level of model detail, and use parameters that generate consistent and coherent media.

R3. Highlighting the Process through Trails. A natural result of working with a non-linear and freeform canvas is a lack of chronological order or direct input-to-output visual connection. For this reason, it is important to have a rich provenance history for each asset exposed a way that allows creators to easily revisit assets and make sense of how they came to be, what inputs created it and what outputs followed, along with the connections between items. These should allow creators to make sense of past decisions, retrace their steps, and explore alternative paths.

R4. Supporting Organization, Collections and Explorability. Because a canvas for thinking and making represents an instantiation of the creative process, it is important for creators to be able to organize and explore information. Creators must have a means to access materials at the right place and the right time. While part of this can be afforded by the spatial arrangements creators may naturally use, they should also be able to search through materials and create relevant collections to easily access important items.

5 Atelier

To operationalize *Protosampling*, we designed Atelier, (Figure 1) where thinking, making, and reflection happen in the same space. Atelier treats the canvas as a living workspace where media are first-class objects that can be arranged freely to be decomposed, sampled, generated, remixed and recombined.

5.1 Overview

The canvas acts as the workspace akin to an artist's studio where materials are brought in, arranged, and manipulated to create new output. We support drag-and-drop upload of images, video, text, audio, and 3D models, as well creating text and sketches directly in-canvas. All media assets can be moved and grouped, then immediately transformed through quick operations that decompose them into reusable components, allowing experimentation and manipulation of materials before moving to more complex processes.

Canvas spatial layouts mean creators can organize information freely, with ad-hoc arrangements developing meaning over time. A mix of collected and generated assets stimulates thinking with engageging around potential triggers for reinterpretation [31].

In physical spaces, interactions often are intimately connected to the setting where they occur [30]. This is especially apparent in an artist's studio, where distinct 'places' hold activity-specific meaning, nearby materials are connected the current task, and materials might migrate over time or cycle back. This inspired our design to localize activity:*quick operations* apply directly on media,

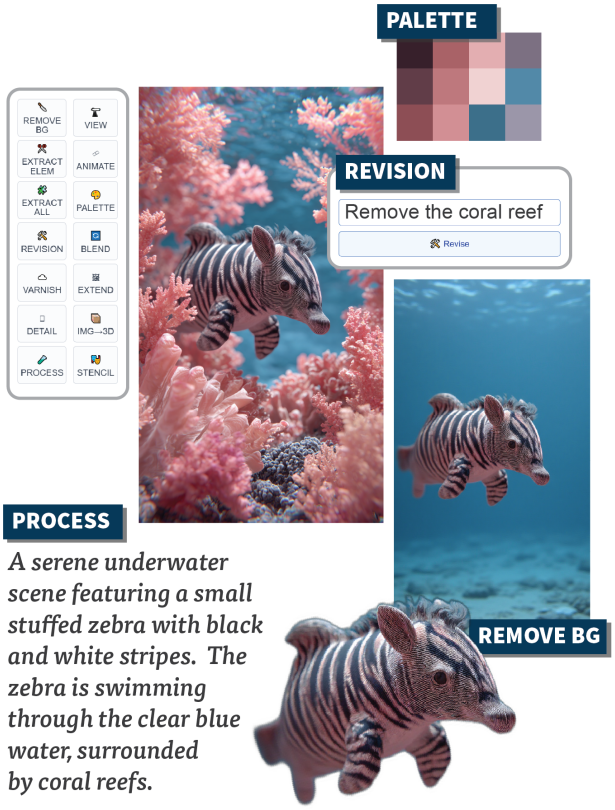

Figure 4: Quick operations. Images in Atelier present common functions for image processing. Illustrated examples include: revision, which allows making edits to the image; process, which analyzes the image, generates ControlNet preprocessors and captions it; and remove background.

and larger operations are done in localized workstations, which we call *Easels*. Easels provide nuanced controls and structure, and the resulting creation is brought directly onto the canvas and treated as potential new material that co-exists with the rest.

Past work often discusses how applications bind work to a given time (e.g., [17]). The state of the workspace thus reflects the lead-up from the process up to that point, its history holding important information. This motivated us to explore means for *organization, sense-making* and *provenance*.

5.2 Quick Operations

Immediately available actions on media are those requiring little to no additional input, such as generating **Quick Sketch** from text, using the contents as a prompt to generate an image. Creators can quickly decompose images into reusable components (Figure 4) with ***Remove Background*** (retains foreground items with transparent background), ***Extract Element*** (extracts targeted elements such as "the blue flower"), ***Palette*** (creates a color palette), and ***Stencil*** (producing structural control images such as depth mask, line art, poses, etc that can be used to control later generations). Creators

can also refine images and make small adjustments with ***Revision*** (single minor revision such as *"give the bunny a mustache"*), ***Upscale*** (creates higher resolution and adds detail), and ***Blend*** (makes style and lighting consistent). Lastly, they can create extensions with ***Extend*** (growing the image beyond the current frame), and ***View*** (generates different perspective views), ***Quick Animate*** (similar to Quick Sketch, generates a 5 second video from the image without additional prompting), and ***Sculpt*** (converts image to a 3D model). 3D models can be manipulated, rotated, and captured as new images that then serve as character and perspective references.

5.3 Easels: Spatial Modular Workstations

Easels are stations where the workspace's materials are gathered and used to generate new media parallel to studio workflows where materials are physically gathered around an area to prepare the act of using them., Within Atelier, assets are moved in proximity to an easel before their use. Each easel encapsulates an act of generation with different goals thus exposing different controls (abstractions of distinct ComfyUI workflows in the backend). These generations are non-destructive: inputs are preserved, can be reused, and iterated on. Easels aim to function as flags for provenance, where an easel's presence and its surrounding materials signal a complex creative intent in that area.

5.4 Easels to Prepare Material

We offer two lightweight easels to help ideate and draft concepts before moving to more complex generations. They provide spaces for gathering, arranging, and externalizing ideas, similar to creating thumbnail sketches before moving onto the final piece.

The ***Collage*** easel (Figure 5) is a station for freeform composition of multiple images, which can then be "glued" together to create a new composite image to be reused. Collage can be used to block out character and object arrangements in a scene. For example, a creator might extract a character from one image, use another generation as the background, adjust size and position, then collage them together. The ***Sketch*** easel allows freehand drawing that can guide future image or video generation. These easels allow for reuse of sketches and collages over time and invites iterative adjustments.

5.5 Easels to Generate Material

Image easels (Draw, Paint, Trace, Modify) contain the workflows for generating new images with abstracted fine levels of control blending text prompts, image references, parameter sliders, and image masks. The Animate easel enables generating videos from images acting as start or end frames.

Draw. The Draw easel (text to image) takes in text as the *positive prompt* of what the creator wants. Optional input(s) include: preset *style options* (Realism, Dreamlight, Anime, Retro Anime, Animated, 3D, Pixel Art), a *details* slider which adjusts the amount of image detail, an *adherence* slider to set how strongly the model adheres to the prompt, a *start image*, and a *preserve* slider to set what percentage of the original image is preserved. The use of a *start image* is akin to bringing in an already existing painting and painting on top of it, where original features may show through, but ultimately a new painting is generated [56].

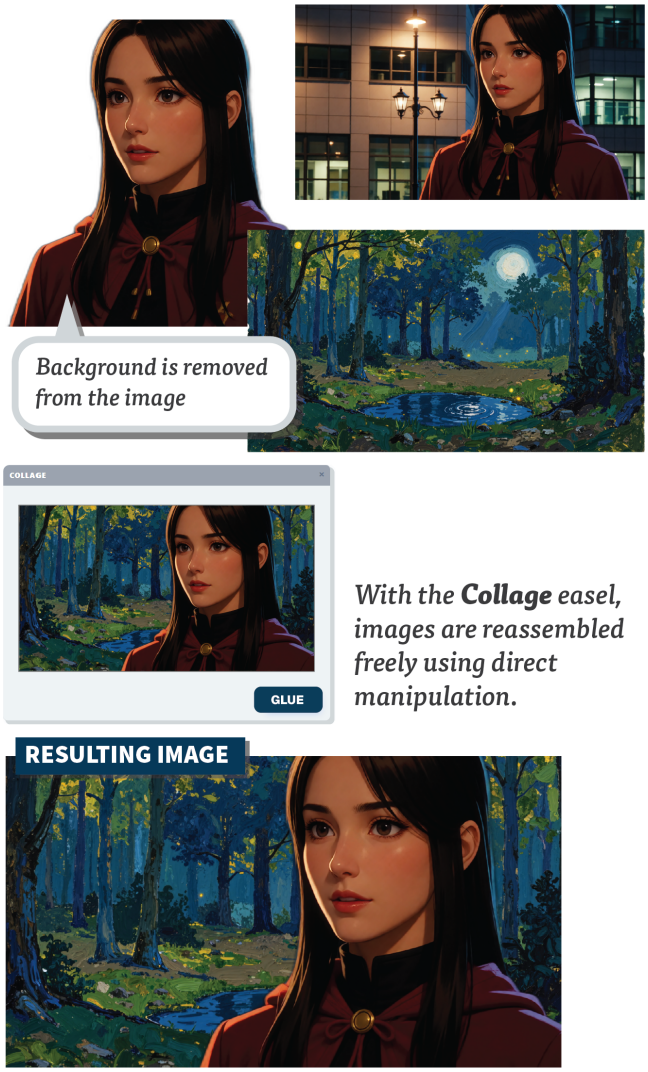

Figure 5: The Collage easel allows composing images into a new image. This example one character extracted from a background and brought into a new scene.

Paint. The Paint easel (Figure 6) extends all the capabilities of *Draw* by enabling up to 3 *image references*, with adjustable strengths of influence. Additionally *reference masks* can be drawn to target areas for those references to be applied. An optional *structure image* chosen from any image on the canvas controls the image layout (depth, line, scribble pose) with adjustable strength. An additional prompt box enables *negative prompts* to specify what is not wanted in addition to the *positive prompt*. The *positive prompt*, *style options*, *details*, *adherence*, and *preserve* sliders are the same as in *Draw*.

Trace. The Trace easel supports generating a new image based on an existing image, selectively redrawing its content and structure. By using two prompts, one to describe the current image and one to describe the desired image, the system can deconstruct an *input image*. Similar to Draw and Trace, sliders for *prompt adherence*, *style*

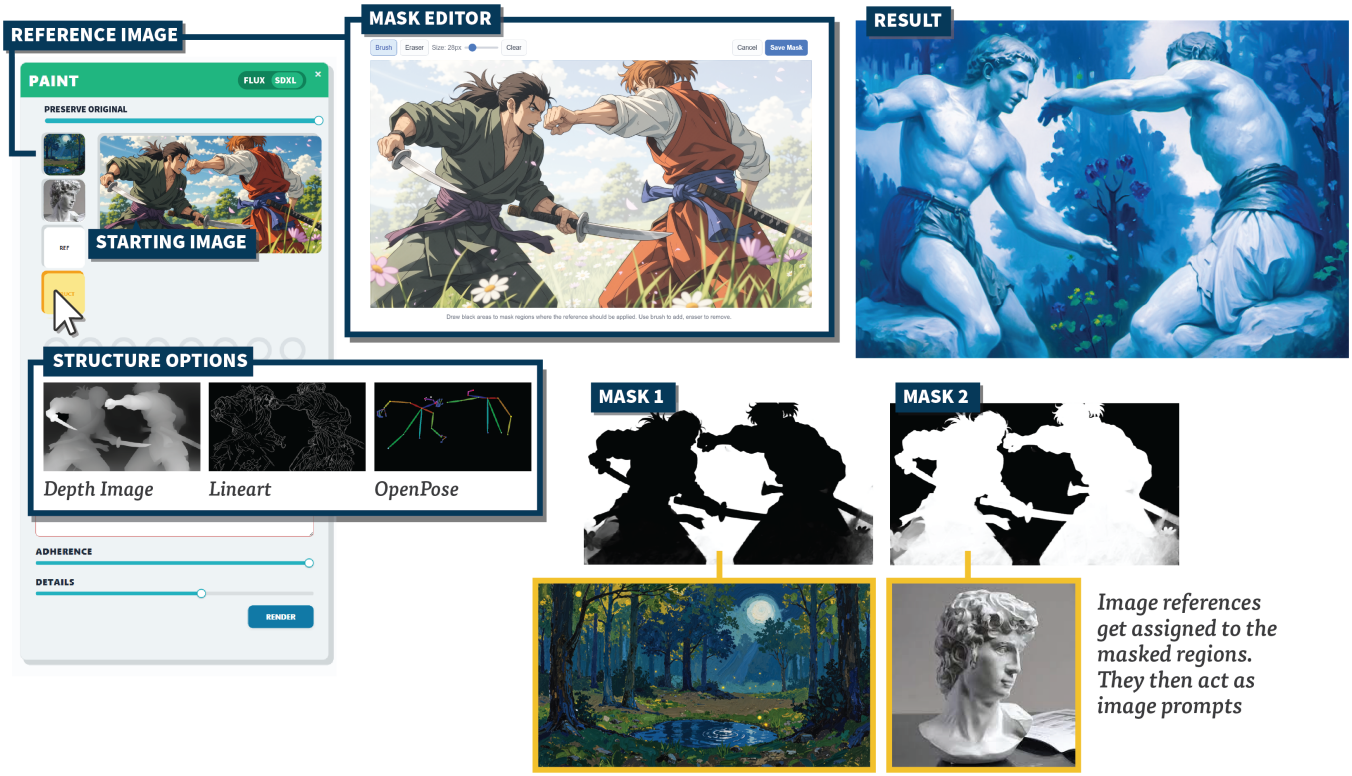

Figure 6: Description of the Paint easel showing how one might add reference images, a starting image and a structure image. The reference image box allows setting the strength and also opening a mask editor to determine the area of influence for that particular reference.

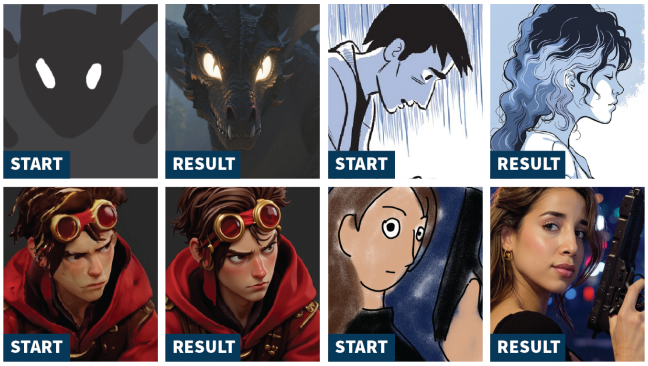

Figure 7: Examples images created with FlowEdit (Trace).

options, *details* and *preserving the original image* are present. An additional slider for the *retracing range* determines at what point in deconstruction to redraw. The easel includes optionally a *structure image* from the input, *style options*, *details*, *preserve* and *adherence* sliders. The trace operation works especially well with coloured drawings, rough 3D models, or for image restyling (Figure 7).

Modify. The Modify easel enables a broad range of edits. To support further discoverability, we extracted a set of cinematographic edits such as changing an image's aspect ratio without major distortions. Atelier provides a set of *relighting* presets)color temperature, direction of light, light intensity, camera angles, image styles). These suggestions, when clicked become additive *prompt pills* (preset text strings that are appended to the prompt), and creators can additionally specify their own custom changes in the prompt. Similar to other easels, *adherence* and *detail* sliders are optionally available.

Animate. The Animate easel creates a video by allowing creators to specify images for the *first frame* and/or *last frame*. The output video creates a smooth transition between the two frames, guided by *positive* and *negative prompts*. Both frames are optional: one can specify a *start frame* without an *end* and vice versa, or do direct text to video generation without any frames. A set of predefined motions in the form of *prompt pills* also allow creators to choose a camera motion or create their own from the example structure. The resulting video is added directly to the canvas, sitting alongside other assets and can be decomposed into still frames.

5.6 Provenance, Organization and Sensemaking

To give structure to a nonlinear and non-chronological freeform canvas, we provide features that support provenance and organization woven directly into the canvas. Provenance in Atelier acts as a living trail, where each asset carries its lineage information

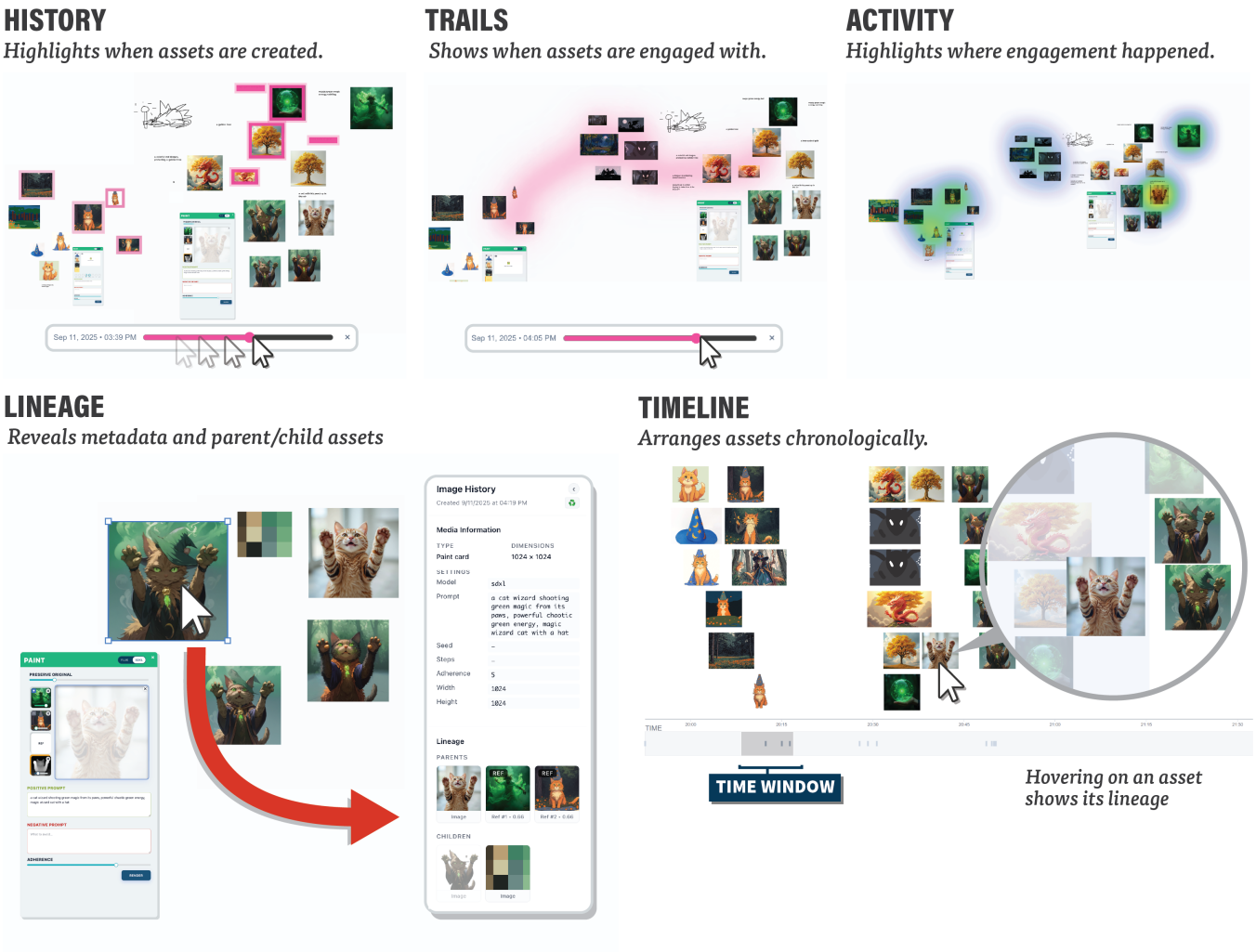

Figure 8: Examples of Provenance in Atelier. Figure shows history, trails, activity, lineage and timeline.

including whether it was imported or generated, and which easel (alongside parameters and assets) generated it (Figure 8).

Lineage. When an asset is selected, the lineage panel shows all relevant information, including the parent assets that generated it and the children assets that came from it, highlighting those in the canvas, allowing for quick navigation to each. From a given image, one can trace back all of the references and parameters that generated it by recursively looking at the parent. At any point in time, they can choose to recreate the easel with the *recreate* button which re-instantiates the generation easel with the exact same parameters, allowing them to work non-destructively from a branching point in the past. Deleted images not present in the canvas still remain recorded in the provenance graph and can be brought back as an active asset.

Asset Emphasis. Outputs might have varied importance, some more useful than others or closer to the creator's intent. Thus, we added an option to de-emphasize assets, reducing their opacity.

History. Examining history features a slider that highlights media on the canvas in chronological order with timestamps, showing the progression of creation across time and space with a sliding window of 5 elements at a time.

Trails. Interaction over time can also be shown through *trails*, a diffused trajectory highlights where in the canvas engagement took place at a given time, aggregating the information shown in history, summarizing the overall path rather than individual elements.

Activity. We also show engagement in a more absolute fashion through a *heatmap* overlay. Time of creation of an asset, the last time it was selected, and the total number of times an asset was clicked are used as a proxy for how much time was spent with it. This overlay uses click count to show the areas of activity, emphasizing the areas and assets creators have spent the most attention.

Timeline. The overall activity over time can be seen in a timeline, which spatially arranges assets on a visual timeline, where hovering on an asset highlights the parent and the children.

Organization. Every image is automatically tagged with an AI generated caption. This allows for a canvas wide search through the caption, prompt and parameter texts to find media through keywords. Assets can also be grouped and automatically arranged into grids. This helps appreciate many materials in close proximity, and also chunking them into a single selection. To support longer term cold storage, we engage **Collections**. Recurring characters, scenes and other groups of media can be saved into collections and tagged. From any other location in the canvas, the creator can pull a copy from these collections to quickly reuse their saved media. These collections prevail throughout the whole project across pages. The images and videos that creators want to save can be added to the **Exhibit** gallery, captioned and rearranged, acting asframes in a film storyboard or final assets for a standalone project. This serves as a space to keep important assets that are specific to the current project. These capabilities support reflection of the creation process, tracking connections between media beyond history, but also as a way to navigate the canvas and find relevant information.

6 Implementation

Atelier was built using tldraw [105], an infinite canvas SDK with built-in text and image features and shape management. Each canvas item is a shape, which can reference external assets (images, videos, etc). We implemented easels, generation logic, provenance, and organization features as custom React components on top of the canvas, using Cursor [8] to assist with implementation. All code was edited and reviewed by the two lead authors. Each AI operation was executed with custom ComfyUI workflows (represented as JSON files). Provenance is represented as a directed acyclic graph in which nodes correspond to canvas items, serving as the source of truth for history.

Images, videos, and 3D models may have multiple copies on the canvas; each copy is a new node pointing back to the original asset, preserving lineage and metadata information from the original source. Parent nodes are media used to create the current child node (reference images, start images). Each node stores key metadata (creation time, last interaction, number of clicks, generation parameters, links to parent and child nodes) also used to create the history and heatmap overlay. The timeline was built using D3[15].

6.1 Workflow Implementations: General Strategies

We tested the Atelier ComfyUI server on computers with RTX 5090 and RTX 4090 graphics cards. These workflows use various local models, including Stable Diffusion XL [4], FLUX-dev and Flux Kontext [35], and Wan 2.2 [111]. Three key strategies apply to many of our workflows, which differentiate them from many currently publicly available workflows. The first two address how images are sampled, and the last deals with multiple permutations within a workflow. For more details on how workflow parameters were each translated into easel interface elements, refer to Appendix Table 3.

6.1.1 Custom Samplers and Lying Sigmas. One novel aspect offered by Atelier's workflows is the ability to add details to images done through Lying Sigmas, a technique that injects noise in the diffusion process[1]. We use the Detail Daemon node[2]. Among the Detail Daemon pack, the Lying Sigmas sampler uses a single parameter for details (dishonesty factor). We choose values between -0.05 to 0. To support Lying Sigmas, workflows need to replace the traditional KSampler[3] with the more complex Custom Sampler, which supports manipulating the guider and the sigmas.

6.1.2 Normalized Attention Guidance. Many AI models offer negative prompts. However, transformer architectures such as FLUX do not provide a negative prompt because they have 'weak guidance' (Classifier-Free Guidance value of 1.0) [44]. A workaround is Normalized Attention Guidance [21], which enables incorporating negative prompts into the system[4], which we set to a value of 9.0 (recommended values are 5.0 to 11.0). This algorithm also enables using negative prompts for models that forego higher guidance for speed-up LoRAs, such as Wan 2.2. Normalized Attention Guidance can be attached to custom samplers.

6.1.3 Dealing with Inactive Nodes. ComfyUI workflows are polled with every new prompt. For certain workflows that take multiple inputs, such as Paint, there are 4 strategies one can take to execute the API: (1) programatically editing JSON files to rewire nodes depending on available inputs (e.g., Paratrouper [58]); (2) creating workflows for each permutation of available inputs; (3) setting parameters to 0 (which is not supported by some features); and lastly, our chosen approach: the use of switches[5]. Switches dynamically reroute nodes in workflows by setting a single number, reducing potential mistakes compared to rewiring.

7 Atelier in Extended First-Use

We conducted a first-use exploration with creative professionals using the system for 4 hours (including breaks), designed as designed as an open-ended session to accommodate learning curves. Our goal was not to evaluate system performance, but to understand how creative professionals might appropriate a tool enacting Protosampling. This merited a study design mixing walkthrough demonstration and observation [57]. Specifically, we wanted to learn how practitioners might use Atelier, different strategies they might adopt, how it might fit existing creative processes, expressiveness, including threshold and ceiling [77]), etc.

7.1 Participants

We recruited 5 creative professionals (2 female, 2 male, 1 gender non-conforming) aged 29-31. We followed a purposeful sampling recruitment [11] and reached out by email to a mixture of direct contacts. Their discipline of expertise ranged in a variety of areas within the media and entertainment industry. Each participant had used generative AI (Table 1) in their creative practice and received the equivalent of a $400 USD gift card as a token of appreciation for participating in the study. The procedure was approved by institutional ethics review boards. Because our aim was not generalizability but rather to surface diverse strategies of usage, 4 hours

[1] https://github.com/muerrilla/sd-webui-detail-daemon/
[2] https://github.com/Jonseed/ComfyUI-Detail-Daemon
[3] https://blenderneko.github.io/ComfyUI-docs/Core%20Nodes/Sampling/KSampler/
[4] https://github.com/ChenDarYen/ComfyUI-NAG
[5] https://github.com/ltdrdata/ComfyUI-Impact-Pack

Table 1: Summary of Study Participants showing their background and experience with AI.

ID	BACKGROUND	AI TOOL EXPERIENCE
P1	**Technical theatre, set design, costume design**	***Current:*** *Claude* ***Past:*** *ChatGPT, Midjourney*
P2	**Cinematography, lighting design, camera operation, 3D, Virtual Production**	***Current:*** *ChatGPT* ***Past:*** *Runway, Midjourney, Google Veo3*
P3	**Digital product design, computer science, music production**	***Current:*** *ChatGPT, Claude, Gemini, Nano Banana* ***Past:*** *Midjourney, ComfyUI, Suno*
P4	**Animation, 2D/3D, stop motion, film, prop making.**	***Current:*** *ComfyUI.* ***Past:*** *ChatGPT, Claude, Leonardo, Midjourney, Suno, Udio*
P5	**Cinematography, animation, visual storytelling, lighting design.**	***Current:*** *ChatGPT* ***Past:*** *Midjourney*

of in-depth rich interaction per participant produced substantial information-rich data.

7.2 Procedure

Participants received an email invitation for the in-person study, provided with our video figure. Participants were asked to think of a project in advance but were not expected prepare extensively. The study consisted of three core activities: pre-study interview, guided demo and free-form exploration, and post-study survey.

Pre-Study: Semi-Structured interview (15-20 mins). After obtaining consent, we conducted *pre-study semi-structured interviews* to learn participants' background, how they use materials, their AI experience and feelings of ownership, etc.

Guided-Demo with Free-Form Exploration (~3.5 hours). The experimenter provided a computer, display and peripherals. Participants used Atelier through a web browser with full access to the experimenter's device, including web browser and creative applications offline to access familiar tools if/when needed.

The experimenter was available for lightweight guidance to reduce friction. Participants led the exploration - features would not be introduced until they had either stopped testing the feature or using past features in combination. The experimenter would ask questions or gather impressions through think-aloud (unprompted).

Participants were introduced to canvas navigation, then shown quick operations. Participants would import materials as they were introduced to different easels, starting with simplest easel 'Draw'. Subsequent order would depend on the individual's project. Lastly, participants were exposed to provenance, organization, and sense-making tools. Participants continued using the system freely for the remaining time. Participants were encouraged to take breaks, most taking 2-3 breaks including one longer 15-20 minute break.

Post-Study Survey (10 mins). After their use of Atelier, participants entered their demographic information and answered

Table 2: Survey responses to paired questionnaire for current AI tools and Atelier. The third column shows the delta between responses for each individual participant.

QUESTION	MY CURRENT AI TOOLS					ATELIER					INDIVIDUAL CHANGE				
	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*	*Strongly Disagree*	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*	P1	P2	P3	P4	P5
Ownership	0	0	2	3	0	0	0	1	2	2	1	1	0	1	0
Agency	0	2	2	1	0	0	0	0	4	1	1	2	1	2	1
Early stage thinking	0	0	0	3	2	0	0	0	3	2	0	0	0	0	0
Late stage thinking	0	1	1	3	0	0	0	0	4	1	2	0	0	1	0
Inspiration	1	0	0	3	1	0	0	1	1	3	1	2	1	0	0
Final Product	1	0	3	0	1	0	0	1	3	1	2	0	1	3	-1
Malleability	0	4	0	1	0	0	0	0	1	4	1	3	2	3	3
Integration to Process	0	1	2	1	1	0	0	0	3	2	0	2	0	1	2
Integration to Tools	0	2	1	1	1	0	0	2	1	2	0	3	0	0	1
End-to-End Support	1	2	1	0	1	0	0	2	1	2	0	3	2	4	2
Reflects Intent	0	2	0	2	1	0	0	0	4	1	-1	2	1	2	0
Enjoyment	0	0	3	1	1	0	0	0	1	4	0	2	2	1	1
Fun	0	2	2	1	0	0	0	0	1	4	1	0	2	1	2
Professional-feeling	0	1	3	1	0	0	0	1	3	1	1	2	1	1	0
Ability to Fix Outputs	0	3	1	1	0	0	0	2	2	1	1	1	1	2	1
Expressiveness	0	1	3	1	0	0	0	1	3	1	2	2	1	0	0
Seen as Tool	0	0	0	3	2	0	0	0	1	4	0	1	1	1	-1
Seen as Collaborator	3	1	0	1	0	3	1	0	0	1	0	0	1	0	0
Empowerment	0	0	4	0	1	0	0	2	2	1	0	0	1	1	0
Creativity	0	0	1	4	0	0	0	0	4	1	0	1	0	0	1

questions about their practice and a set of paired Likert-scale questions contrasting their current use of AI tools with Atelier. An example question is: *"I feel a sense of ownership with what I create with [Atelier | my current AI tools]"*, see Appendix C for full details.

7.3 Data Analysis

All sessions were screen and audio recorded with participant consent, then transcribed with filler words removed. The qualitative data was analyzed jointly between the first and second author, the

first author conducting open-coding on the transcripts. The second author, who conducted the experiments, reviewed the codes and data. Through iterative discussion and thematic clustering [20], the authors identified a set of themes. Post-study survey Likert-scale data was used to complement the findings and analyze shifts in sentiment across individual participants.

8 Results

Participants created canvases for their indivdual projects (Figure 9). P01 worked to materialize a "brick made of clothing scraps" for set design work. P02 created a set of film scenes of a man playing a guitar outside of a casino, iterating on lighting, camera angles, and camera movement. P03, after open-ended exploration returned to an old app idea that rendered stock graphs as mountain ranges, using Atelier to prototype the visual translation. P04 recreated an old animation project of numbers morphing into different shapes as a movie countdown. P05 explored shots for a film concept and then built out product shots for an advertisement project, iterating on light, framing, and movement.

8.1 How was Atelier used?

Favourite Easels. Participants all had favorite easels. P1 and P4 preferred Paint. P1 liked the negative prompt and structural reference as it enabled them to have more control of the composition, whereas P4 noted that it gave the most flexibility and options to explore. P3 and P5 preferred Trace because it allowed them transform starting images: *"I have this 'retracing steps' as a control mechanism for adjusting the severity in a bit more of a predictable way... I could become a power user of [this easel]... This gave me more predictable results"* (P3). P2 saw the strongest affinity for Animate: *"the ability to see it in video, see it in motion gives me so much to think and to reflect creatively... so much as a possibility [compared to] an image [as] extra step of my creative process... it changes everything"*.

Biggest Challenges. P1 and P2 described Trace as the most challenging easel. P2 was unclear on how it worked despite having achieved some successes. Having to describe the original image and then the final image was something that all participants were initially confused about. Across all participants, prompting was a major challenge. P1 and P5 resorted to using Claude for support - P1 would ask for better phrasing of a prompt, whereas P5 would use it to find the right description words or to translate from their native language to English. P2 noted with generative AI, often the only way to obtain variations is to *"write better, just give a better prompt... it's very limited of what you can do"*.

Clarifying Design Intent. Atelier provided participants with more ways to achieve the visuals they wanted. Partly due to experimentation, with P2 "lik[ing] how precise [one] can be with the program... options of Draw, Paint, Modify gives you more room to experiment". P1 described how: *"it gives you control... it's easier to know how to imagine... this is closer to what I want"*. Creating different images often served to give a better sense of the design intent: *"it's nice to see other explorations... Now I am liking more this idea than [a previous idea]" (P5)*. Undesired results also sharpened design intent by clarifying what they did not want, reducint the space of possibilities: *"it would give me clarity... I don't want the patchwork brick, I want the scrapped brick. So then I'm gonna shift how I'm gonna build things so it can help me think, having clarity of what I don't want. It's faster not to get lost"* (P1).

Support for Different Stages of the Creative Process. The post-study responses suggest that participants can envision using Atelier for inspiration, early-stage thinking, late-stage thinking, and final products more than their current AI tools (Table 2). However, perspectives depended on the practitioner's primary medium and their goals, and was primarily seen as a conceptual tool. P1, despite having used AI images for a final project in the past, feels that generative AI is conceptual because it remains digital, whereas set designing is a physical hands-on activity: *"I would use images to visually communicate. Take the concepts, take the ideas, take my interest, and communicate and then arrive at a thing that people are like 'yes, that's what we're building'"*. P2 noted how Atelier would enable them to create more realistic pitches: *"I see [myself] using [it] for pre-visualization from every stage... thinking about things and brainstorming... reference videos or have reference images for a pitch package or a grant application... pitching to a director with this saying, 'oh, this is my vision, this is how I thought about your scene'* and being able to show specific types of movement and tone, where "*it's kind of hard to beat when you show something that is unique versus an image of a movie that already exists.*" P2 also said they might use Atelier to create full videos for fun. P3 saw value in concept generation, but also envisioned their easels as pipelines they wished they use could reuse: *"what I would love to do is create this pipeline and then ultimately export it or serve it as an API for my own apps"*. P4 and P5 saw value in using Atelier for storyboarding or animatics, P5 especially noted it being useful for *"commercial stuff... they really want to see how it's gonna look. And if you're bringing something like this, you're going to save a lot of problems with them during the shoot"*.

8.2 How did Participants Organize the Space?

Canvas Layout. Participants developed their own ways of moving through the canvas. Each participant had an element of working left to right. P2 and P3 organized their canvas straightforwardly left to right, with P2 explaining: *"I tried to think chronologically... moving the way that we read in the Western world... trying to keep that consistent"*, noting that if they were working on a longer-term project they would create separate corners for each easel. P3 moved left to right temporally and arranged variations vertically. P4 kept outputs on the right and references on the left closely to the easels, noting: *"If I was organizing it with more time, I would be a bit more separated into more sparse clusters so that I have more space around it to work"*. P1 moved from the center outwards, making preferred work bigger placed to the far right, with discarded material on the far left. P5 moved from left to right, then started a new row below the current work when exploring a new topic. Most participants (P2, P3, P4, P5) clustered outputs closely around the easels that generated them, with P4 noting that even when zoomed out: *"I can visually tell what's what... little cluster over here, little cluster over here... I can tell different things going on."* Participants took different approaches to managing volume of generations. Most (P1-P4)

Figure 9: Canvases created by participants using Atelier. P1 shows a concept for a brick made out of fabric scraps; P2 shows shots for a scene of an old man playing a guitar inside a truck; P03 created a variety of concepts, including an app for visualizing stock trading as landscapes; P4 recreated an old animation exercise of numbers made of different materials morphing; P5 explored different shot concepts for a film and then for a commercial.

kept the majority of their generations, deleting only outputs that were completely off-target, whereas P5 deleted both easels and generations more liberally. Several participants (P2, P3, P4) used the grouping and packing tools to spatially lay out their work, while P1 and P5 were more comfortable with canvas elements overlapping.

Easels as Spatial Anchors. Easels functioned as anchor points within the infinite canvas. P1 articulated that with *"canvas based work, it's so huge, you can get so lost. I think the easels provide you some structure without limiting you... it helps you to organize things... so you can work within the flow happening, but then you can also organize so that when you're back, you don't get overwhelmed by the amount of stuff"*. They explained this would allow them to come back to work on a new day with a *"fresh place to work while still being able to draw connections and go back on the work"*. Participants valued the space of the infinite canvas with P2 describing: *"any canvas style creation, it's almost like an infinite. It's a multiverse. You can create a multiverse of ideas and options... I like how it feels endless"*. They noted that even with more erratic organization, they *"could still see the process"* spatially. Participants sometimes treated easels as personalized workspaces they configured for their own workflows. P3 described easels as: *"I see this as my little factory area... my playground... this is my space... there's so many different parameters and ways to set things up. I don't want to even slightly forget how I did it. I like to keep that persistence. It's almost like I built the tool because I added all these parameters in it"*. P2 envisioned organizing the space by easels with extended use: *"if I was gonna work this more seriously, I would probably separate a bit more tools. This is my Drawing corner. This is my Modify corner"*.

Provenance Tools. Other provenance tools such as timeline, history, heat map, lineage panel and recreating easels options were used to navigate and make sense of their work and the process it took to get them there. P2 called the history feature *"life changing... why doesn't everything have this, like a history, a visual history of what you're doing"*. P3 explained *"I like to preserve the history of what I've done. It both makes me feel like I'm making progress and I can see where I've come from, revisiting old ideas when maybe I've forgotten about them. I think that's pretty important"*. The "recreate easel" feature was used to return to specific parameter configurations with P1 describing it as *"the gift of being given your image and then the notebook that shows you how you got to the image... you can go back to your sketch of the painting easily"*. P4 valued how the provenance system addressed reference management challenges noting it is *"really useful. It's good because then it safeguards your process. It's dummy proof... the management references... you [might] have to go to an outside software... but then you have it within the software here, which is handy. And then you make sure that you don't lose it because it's in your project"*. The preservation of their process was important to participants both for managing references as well as for making creative connections and reflections.

8.3 How Does Atelier Compare to Other AI Tools?

When using Atelier, participants brought up comparisons to previous experiences with AI image or video generation platforms, including prompt-based platforms such as ChatGPT and Midjourney, as well as node-based platforms like ComfyUI.

Comparison to Prompt-Based Interfaces. Compared to prompt-based interfaces, Atelier's multi-modality makes model capabilities visible through the easel's design (e.g., slots that can be filled for image references) (Table 2). To participants, language-based approaches do not provide clear means to create with generative AI: *"[one has] to communicate to a computer only through words, which is limited to represent what [one has] in [their] brain. Instead of allowing the artist a little bit of autonomy with the tool, it makes [one] go through that [language] layer... how can you represent what you have in your mind physically if you don't have all the tools available to you, if you don't have a hammer, if you don't have a chisel..."* (P4).

Compared to prompt-driven interactions, Atelier shifts participants' relationship to the material from talking to a model to manipulating content: *"When I go in Midjourney or Runway, I feel like I'm interacting with the AI... I'm asking something from it and receiving that image. Here, I feel like I have creative control. It's my creative workspace and not the AI's workspace"* (P2). P5 noted that in comparison to Midjourney, Atelier provided starting points: "*all these options like Draw, Paint, Modify... feel more friendly. You have more options to start... that's the most intimidating part for me when I am trying to create something"*. This shift gave participants agency when generations failed, rather than blindly rewriting prompts, they had *"accessible knobs"* to adjust.

P2 felt that using Atelier brought creativity to the process itself: *"it's a creative process even to interact with the system. While when I'm working with other video generative tools, it doesn't feel like it's creative"*. P2 suggested that many details, from the vibrant colors making the system feel approachable, to the naming conventions in Atelier encouraged creativity: *"what I feel is creative is the tools here. Even the way you're naming the tools, like drawing, painting, sketching, collage... it's mentioning art forms, and it's encouraging me to think in art in a more creative way rather than prompting*".

Comparison to Node-Based Systems. P3 and P4 both had prior experience with ComfyUI. They articulated that Atelier was more accessible and intuitive to learn and use. While node-based systems offered more fine-grained control, the learning curve served as an intimidating barrier to creation, where many participants saw the software was a means to an end and if the *"learning curve to the software is as big as the art I'm gonna make, I don't want to do it"* (P1). P3 recounted that tools like ComfyUI: *"are very daunting tools.. the different inputs and outputs, little connectors and knowing where they go and how they operate... there's just so many... even coming from the software development world and using Blender, I haven't fully adapted to nodes... I tried [ComfyUI]... and I failed because it was too hard of a learning curve"*. In contrast, Atelier: *"at least allows me to experiment with these pipelines in a bit more of a novice way"*. P4 noted that Atelier *"feels a lot more personal... more collaborative...*

more visual, it feels easier to explain, feels easier for even somebody who perhaps doesn't have experience using tools... visually they can already understand something, versus with ComfyUI you have all this text and nodes... and then finally an image at the end. With [Atelier], it's only images. It's easier to show this [easel] is generating all these images... versus a whole system generating one image". This made the work legible for participants' own understanding and for potential collaborative sharing. P1, often working with less computer-literate collaborators, noted that *"there is something about this that makes it easy for people to use that you don't need a lot of buy-in on how this could be useful. You get the results as you're working through it"*.

The increased speed at which participants could get desired results allowed for context to be maintained during creative exploration. With all the generations in one place, Atelier provided a continuous visual history, which P4 referred to as a *"fresh memory"* of how ideas developed. Moreover, because the iterations were close together in time and space, they could *"trace it back mentally"* and *"see it all there"* by zooming out. P4 noted that if they used ComfyUI, their generations would have required them to split the work across two days. This continuity enabled forms of comparison and judgment that were difficult in chat-based prompting, which P3 described as *"rigid"* and *"utilitarian"*, oriented toward reaching a single output rather than exploring alternatives: *"this tool allows me to basically explore a search space of creative possibilities... whereas other AI tools I've used don't really use this canvas approach... feels more task focused"*.

8.4 How does Atelier Support Craft and Ownership?

The pre-study interview highlighted some base beliefs from participants regarding their views on craft and ownership. Yet the use of Atelier also highlighted new ways to reflect people's craft and ownership. Specifically, we identified themes of intent, malleability, collective outputs, and the participants' ability to understand AI as a medium. The post-study survey results (Table 2) also highlight elements relevant to these changes, namely the categories of: ownership, agency, malleability, reflecting intent, ability to fix outputs, expressiveness, empowerment and creativity. All of these categories saw mostly sentiments in favor of Atelier.

Intent. One way in which participants felt ownership is through intent. In the pre-study, P2 described how their generations are often unique because they *"have very strict ideas [their] mind and they take so many different prompts"*, while P3 described how in the past they made an AI comic where they felt they were *"creating that art form"* because they were *"being very intentional"*. The usage of Atelier also demonstrated how intent was reflected. P1 described: *"It feels like something I had in my head before and AI just made... the image generation is just a tool... to making sure that people understand and visualize the same thing as I visualize"*. P5, while unsure about their sense of ownership, felt Atelier enabled experimenting to refine their design intent: *"this thing already has some ideas that I have or a mix of ideas that I want to reach*".

Malleability. Malleability showed the most considerable increase in the post-study survey. In the pre-study interview, P1 described how they need to *"have a lot of input into [AI] to then feel like it can be [theirs]"*. P2 described how prompts are achieved from many iterations: *"I crafted them like a diamond... the diamond was always there but then it needs to have that cute shape. Otherwise it's just a rock."* These points highlight how each generation leads to adjustments in parameters for rapid iteration. After using Atelier, P2 found more nuance in how they see ownership *"here I feel like I'm a creator. I'm creating, combining, sketching, drawing... it's about the process... prompting feels a little more technical... You're trying to guess formulas and what the AI thinks and try to break those formulas"*. P3 noted how easels embodied a set of parameters that enable manifesting their design intent: *"[in] getting [the concept] accurate and right... the pipeline to achieve this technological possibility ability is that there was craft in that."* This sentiment was echoed by P4, given how in tools like ChatGPT *"you have to write a whole paragraph of what you really want. But then that might not be the result"*. In contrast, *"with [Atelier] you isolate all those characteristics and then you can make sure that all those aspects stay the same even after you modify other things."*.

The Collective Outputs. The malleability aspect considers how generations lead to new settings and iterations. This process can also be looked at from the point of view of the outputs generated which over time, can reflect changes in settings as generations went in specific directions. P2 had already described the iteration needed to get the right prompt. However, since Atelier places all outputs within the canvas, P3 noted a sense of craft and ownership from seeing the different generations side by side, reflecting on the spatial arrangement and the emergent results - their sense of ownership and craft came from seeing the journey: *"even though the source material is from a very famous film, as we were going along... picking up different variations and worlds... I could start to pick and almost make opinions about what direction I wanted the style to go... And I think part of it is a having the control to guide things much easier... But also seeing the variations, seeing them all laid out in this way, being able to accept and reject and ultimately pick a creative direction from a large set of explorations... is part of that creative process that gives me ownership"*.

AI as a Medium. When discussing why they do not feel ownership over their AI generations, P5 explained it was because they do not understand how generative AI works, in particular: *"It's a challenge sometimes... I get desperate really easily. Like, that was not my idea..."*. P5's use of Atelier appeared to create a sense of control and agency, especially in their use of the Trace easel, enabling them to generate similar shots to their reference material but better aligning to what they were looking to create. While the feeling of ownership did not change, P5 developed more confidence in their experimentation, also reflected in their post-study survey, with increases in agency and ability to fix outputs. P4 described similar importance to understanding AI as a medium: *"It's like accepting what it is and using what it can give us to shape what your imagination is trying to do... it's similar to what artists do physically too... at the end of the day, what you imagine is not necessarily what you end up making.*

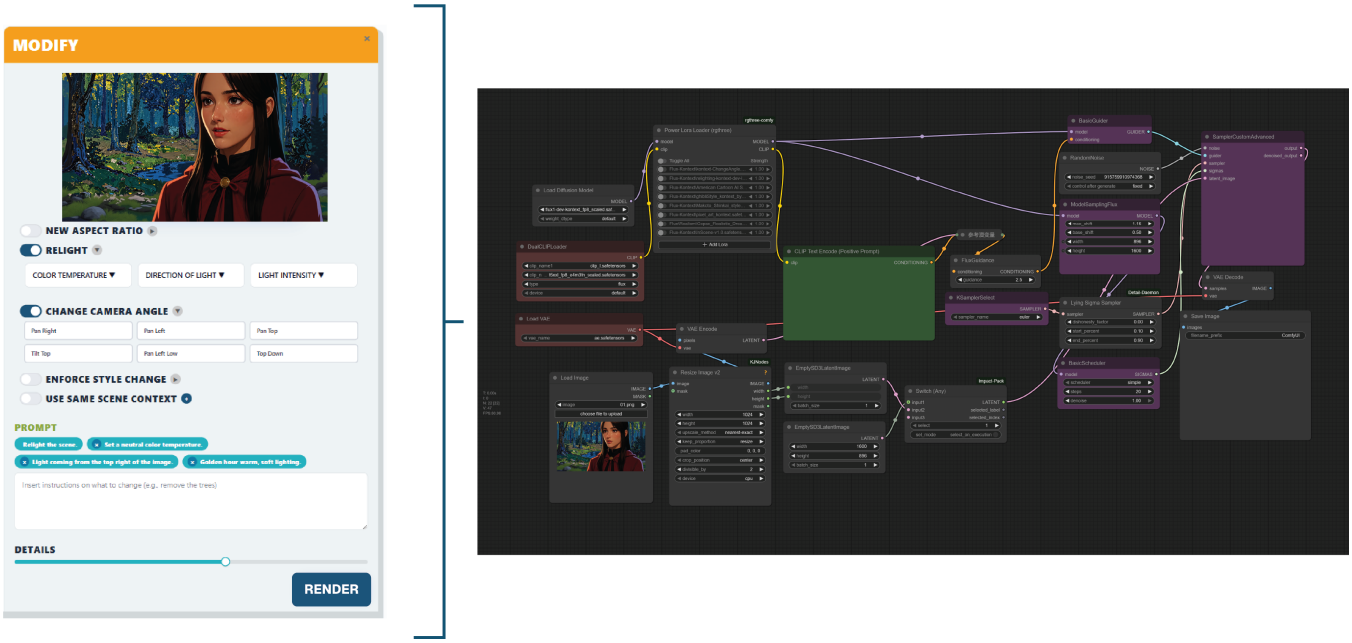

Figure 10: Example of the Modify easel and how it maps to its corresponding ComfyUI Workflow.

There's a lot of things in the path of getting something done from beginning to end that can change". While we did not explicitly ask about understanding as a medium, the sentiment responses to the post-study survey (Table 2) show a change in a few metrics that could be associated to understanding AI as a medium: malleability, agency, ability to fix undesired outputs and sense of empowerment.

9 Discussion

Protosampling as a concept provides a lens with which to look at the creative process in order to inform the design of authoring systems. Atelier represents one such way of doing this, acting as a validation of the concept [115]. As part of the research process, using Atelier instigated critical reflections as a means to Protosampling, the challenges and opportunities of generative AI being non-deterministic, to the conceptual implications, technical and conceptual limitations, and potential for future research.

9.1 Atelier as a Means for Protosampling

Early in this paper, we proposed Protosampling as a method that balances the curatorial nature of sampling and its extraction of insights, with the act of prototyping which manifests ideas.

9.1.1 Blending Thinking and Making. The media-first approach mirrors the real world [30], where references and inspiration are dealt with as wholes since the creator may not know why something is useful yet [52, 64, 65]. Easels, together with the quick operations, act as the key way to enable the 'making' part of the rationale, whereas using a canvas, and provenance and organization techniques are more geared towards 'thinking'. Many of our participants described that the interface was very intuitive, where the space felt like it could be called theirs. We also found that Atelier directly supports the procedural actions in Protosampling, particularly derivation and combination approaches. Aspects such as analogy, metaphor, and first-principles are processes that are much more complex and more up to the practitioner to devise as part of their creative problem solving, where creators might find more value directly engaging with other people or learning more about the world around them. Sampling and prototyping were not treated as mutually exclusive operations, but rather interleaved within each participant session. For example, a style transfer might have functioned as sampling when used to survey possibilities across variations, and at the same time functioned as prototyping when its output became a substrate for further generations. Importantly, every participant emphasized that working in Atelier felt playful or fun. This sense of playfulness appeared alongside behaviours such as freely branching, recombining, and testing ideas, suggesting that the Atelier supported exploration that felt open and low-stakes.

9.1.2 Encapsulation. With current approaches, creators have to decide between two extremes. One is to use a dedicated service, such as Midjourney [71] or Runway [90], with the interface and options provided. These can be fairly accessible, though the scope of each service varies, where work often has to migrate across services. The other approach is to use systems where they can access the models directly such as ComfyUI or a gradio-based interface such as Automatic1111[6]. However, the challenge, as shown with participants who felt frustration with ComfyUI or node-based canvases which take time and training to learn, becomes how to harness the potential of these node-based interfaces and make them more accessible. This is where the potential for designing interactions comes in. These challenges motivated mapping Draw and Paint

[6]https://github.com/AUTOMATIC1111/stable-diffusion-webui

to enhanced text-to-image and image-to-image workflows, Trace to use FlowEdit, and Modify to use editing models such as FLUX Kontext. The selection of model, checkpoints, LoRAs, etc. came second and is prepackaged.

Early in ideation, preprocessed structure images initially lived in the canvas with other media but soon flooded the canvas and required users to think about how to use them. Moving them to become suggestions in easels that had a structure compartment removed complexity while retaining their usefulness.

One example of how easels maps to a ComfyUI workflow is shown with the ***Modify*** easel (Figure 10). Besides using FLUX Kontext-dev and LoRAs for relighting, camera changes, and styles, we extracted potential keywords from the different LoRAs and added them as prompt pills. The toggles on the interface map to different LoRAs or to the switch for aspect ratio.

These examples illustrate the many considerations in building abstractions, looking to have low thresholds and high ceilings [77], provide a set of building blocks [57], and expose the right parameters so that expressiveness can match the task as much as possible [80, 86]. In building these abstractions, we considered experiences from different online communities from daily engagements such as Reddit[7] and Banodoco[8], as well as existing documented first-person accounts [56] and our own experiences learning and using these technologies. P3 stated that after using Atelier for the period of the study, they felt that they understood ComfyUI more. This points at a possibility of progressively introducing more advanced concepts or unlocking more complex features in the system, and could further explore creative systems as learning grounds for AI concepts through use.

9.1.3 Organization and Provenance. From participant use, content often congregated around easels, acting as potential anchors for content. At the same time, the canvas metaphor makes it so content is scattered without structures, even without piling as an option. Given the discussions on knowledge work, storage and search, and material retrieval [13, 14, 42, 97], we participants saw value in being able to explore the content and create more permanent collections. We believe that engaging with these externalizations can invite reflection [29, 95], as now the entire set of materials exist as triggers for reinterpretation [31]. This suggests that media-first and freeform creation can be viable, but an infinite canvas needs to be paired with equally powerful sensemaking features in order fer participants not to get lost. Creative work is often very collaborative, and participants had others in mind that they wanted to show the work, whether clients or collaborators. The space lends itself well to collaboration or sharing of process, as participants felt that it was easy to understand their own journey.

9.2 The Two Sides of the Coin of Open-Endedness and Activities

A challenge with generative AI is its lack of determinism. On one hand, this can be a powerful advantage, offering many ways to solving a problem. For example, to deal with the challenge of creating consistent characters, one can use the **Modify** easel to create alternative camera angles for the character. An alternative is to use **Animate** to generate a video from the character image, then extract still frames. Together with the ability to remove backgrounds and reassemble in a collage means that it is possible to achieve multiple character poses and positions. In the user study, we also found participants using easels and features in ways that we did not expect. For example, P1 created multiple copies of a brick and arranged them in the cavnas as a way to think about building with physical bricks, while P2 used Collage to create composite shots that could be brought into Trace. Collage was especially useful when other easels were not following their exact intent, such as rendering the wrong style of truck.

This open-endedness can also be a challenge, especially when the limits and capabilities of a model are not fully understood. Terminology such as 'denoise' or 'schedulers' can be difficult for novices to pick up, and these are terms that exist outside of most creative fields. While new terms are a part of any tool, for example, how does one learn the best time to use FlowEdit versus Flux Kontext? Both can restyle an image, but FlowEdit can provide more creative freedom. Even with more experience, the answer is not straightforward and invites multiple tries and learning from failures. However, this also leaves opportunities for surprise even when understanding how to use these techniques.

9.3 Conceptual Implications

While the primary concern within Protosampling is the focus on procedural action and the trail, the reflection resulting from this activity can have a direct impact on the creative process itself.

Transformation. The first conceptual implication of Protosampling refers to the transformational nature of a reflective practice [94]. Evaluating what has been collected or created by a certain point will inevitably reframe the creative problem. Ohlsson [79] claims this can take place by elaboration (representation changes from adding information), re-encoding (rejecting part of the original interpretation of the problem and revising it), or constraint relaxation (changing an incorrect representation of the goal).

Emergence. The second conceptual implication is the emergent nature of the creative process, which Cross describes as unrecognized properties that are integrated into future concepts [26]. This highlights the interplay between being intention-bound (aligning to original plans) and emergence-driven (openness to external influences, learnings and inspiration) [37]. Ideas do not exist in isolation [27], instead they are all part of a connected, directed and rational process [113, 114], one in which externalizations continue to invite new follow-on ideas [93].

Bias and Fixation. An interesting question is whether the use of generative AI can increase bias or design fixation. All of the models have different biases and optimizations. With Atelier we support a variety of models and activities so that practitioners can try different ways of solving a problem. By having reference imagery the models are guided more towards the reference and away from its base training, same with using individual styles. In our experience having generated over 100,000 images with local AI, we also found that the different models and techniques have different 'qwerks'. For instance, Stable Diffusion may have issues rendering

[7] https://www.reddit.com/r/StableDiffusion/
[8] https://banodoco.ai/

fingers, but its inability to fully follow the prompt enables blending multiple styles. With FLUX Redux, reference images have features extracted that lead to unexpected recombinations. FlowEdit, used in Trace, adds nuance of how much distortion to create from the starting prompt to the ending prompt. Models like Wan are better at prompt adherence and consistency, but can often accidentally infringe on intellectual property if the prompt is vague.

One way in which design fixation can come up is when someone engages with more iterative refinement over exploration. During the study, P1 often talked about following a certain path and continuing to refine and feeling distracted: *"it can be distracting because it can generate so much stuff that then you're like, oh, is this what I wanted? Or is this just exciting?*". With more physical activities, a clear limit is fatigue, but with AI generations that can go unnoticed for a long time.

9.4 Local Models and Implications

A key design choice in Atelier was to rely on local, open-source models rather than cloud services. While cloud services offer many benefits, such as not having to rely on individual hardware, much of the controllability and nuance that makes Protosampling possible is due to open source technologies. Open source models allow access all the necessary parameters to control them, making it possible to mix and match different platforms and extensions, such as the ability to add details or negative prompts or trigger multiple styles. For example, a Paint easel can generate an image with the base model and then one can tweak it with a realism style with 30% strength, a dream-like style with 80% strength and a retro anime style with a 60% style[9]. This flexibility that enables combining different models for image referencing – with Flux Redux[35] the model automatically extracts items from each reference depending on the strength setting, whereas with USO[116] the style is explicitly isolated. Other arguments for running AI locally include energy efficiency, preserving privacy and avoiding content trapped in walled gardens, and the ability to choose which models to work with. In this version of the system we preselected style models, but one can imagine the ability to import one's own finetune of a specific character or style. This brings a whole realm of possibilities for creators to truly harness the power of generative AI for creation.

Model flexibility is also where ethical concerns come into play. AI models were trained using publicly available data without opt-in or opt-out options, and the economic impact to the creative community has been noticeable. The ethical implications are complex and still under active discussion, since there is nothing stopping bad actors from creating malicious or inappropriate content or training on the work of specific styles when working locally.

9.5 Limitations

The current design and implementation of Atelier has three main limitations. Technically, while ComfyUI is a very powerful tool for prototyping and has the benefits of being local and supports open-source models, it is not designed for robustness. Challenges include paths and models that are often hard-coded and memory leaks, and with so much driven by community contributions, occasional updates can break the system. Moreover, we are running the system locally on consumer hardware. While speeds can be comparable to commercial-grade hardware, it is not possible to hold all models in memory. This means that models are constantly being swapped out, increasing wait times for generation. These limitations are not inherent to Atelier's interaction model, and as model-serving infrastructures mature, the same workflows could be deployed with significantly greater robustness and usability.

Secondly, Atelier runs on the assumption that the bulk of the work is done in the system. While one can import and export much of the media, Atelier becomes another place for information to be potentially fragmented across applications. This is where a system like Atelier might work better if the metaphor could be applied to the operating system as a whole, similar to how it is done in some other systems [17]. That said, study participants (P2, P4) described wanting to stay in the tool, and wanting to avoid switching applications even for things like image editing for which other tools might be best suited.

The third potential limitation is conceptual: while we argue for references and generated assets to co-exist in the same virtual space, success with generative AI requires a large number of generations [56], leading a large portion of what gets generated to effectively be scrap material. The other potential challenge with having so many generations is it may be easier to lose sight of what one wants to create, whether due to compromise on a model's limitations, or simply having so much content saturating the mind and forcing incidental emergence. AI models also have a defined set of biases in their data and training, and those can also get perpetuated. While having multimodal inputs as done in Atelier can help with this, there is still a risk that solutions get pushed towards these biases, potentially leading to design fixation, or accidental infringement on other intellectual property. These are simply realities of dealing with AI as a medium in this current time, and will require more time and studies to better understand and address.

Additionally, the evaluation of Atelier was conducted with a small sample of five creative practitioners working on highly personalized tasks without a baseline. Rather than serving as a comprehensive evaluation, this serves to contribute qualitative patterns on how practitioners explore, structure, and reflect on their creative process with Protosampling foregrounded in the workspace.

9.6 Future Work

The work carried out in Atelier opens a set of interesting research opportunities. The first one is the exploration of video, given that thus far we only explored image-to-video. There are many more ways to create generative video that have even more controllability [56], and those modalities need appropriate abstractions and gaining a better understanding of the interactions at play. The canvas approach for video invites thinking about abstractions, but also reflects on how there might be potential interplays between the physical space and the time-based nature of animated motion.

Another potential area is exploring how to make more technically robust infrastructures accessible. For example, Atelier currently supports a small set of LoRAs and checkpoints, yet being able to have specifically trained characters could dramatically improve the different activities supported by the easels. There are also

[9] strengths do not need to add up to 100%, they just imply how much of the individual finetune gets embedded into the main checkpoint

questions for what it means to add more intelligence to the canvas, or support chaining or nesting workflows in a way that does not require computational thinking.

Lastly, exploring protosampling and operationalizing it invites research into having a better understanding of how practitioners might use Atelier for their own creative process, and seeing how their ideas and thoughts evolve over time. Interesting future work could involve practitioners using the system over a period of time to see what they can create, what strategies they devise, and how different people make use of the canvas and the tools provided.

10 Conclusion

Sawyer's compilation of creative theories show how insight, no matter how sudden it may seem, is a systematic process always connected to a practitioner's prior work, their sketches, notebooks, etc. [94]. This highlights the importance of curating collections, manifesting ideas, and reflecting on the process. Tightly interweaving sampling and prototyping – *Protosampling*– helps better understand the relationship between the collected information and how ideas get manifested. This is especially true now that generative AI has accelerated and blurred the boundaries between these two important activities. Deriving and operationalizing Protosampling, and understanding how information moves about in the process enabled designing Atelier as one open-ended way to blend thinking and making, abstract complex workflows into easels as self-contained modules, and make sense of and organize the work.

Atelier was built with Protosampling at its core. Atelier looks to make the creative process more tangible, and speaks to how creative work in this new age of generative AI is not about finding the right prompt, or accessing the latest model, but by the small decisions and the deliberate actions creators take, and how the right tools can empower them to make their ideas come true.

References

[1] 2005. *How do you design?* https://www.dubberly.com/articles/how-do-you-design.html

[2] 2026. *r/StableDiffusion.* https://www.reddit.com/r/StableDiffusion/

[3] Adobe. 2025. *Adobe Firefly.* https://firefly.adobe.com/

[4] Stability AI. 2024. *Stability AI.* https://stability.ai/

[5] Alibaba. 2024. *Flux Turbo.* https://huggingface.co/alimama-creative/FLUX.1-Turbo-Alpha

[6] Rifat Mehreen Amin, Oliver Hans Kühle, Daniel Buschek, and Andreas Butz. 2025. Composable Prompting Workspaces for Creative Writing: Exploration and Iteration Using Dynamic Widgets. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25).* Association for Computing Machinery, New York, NY, USA, Article 144, 11 pages. https://doi.org/10.1145/3706599.3720243

[7] Victor Nikhil Antony and Chien-Ming Huang. 2024. ID.8: Co-Creating Visual Stories with Generative AI. *ACM Trans. Interact. Intell. Syst.* 14, 3, Article 20 (Aug. 2024), 29 pages. https://doi.org/10.1145/3672277

[8] Anysphere. 2025. *Cursor.* https://cursor.com/home

[9] Silvano Arieti. 1976. Creativity: The magic synthesis. (1976).

[10] Banodoco. 2026. *Banodoco.* https://www.banodoco.ai/

[11] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. *Qualitative HCI research: Going behind the scenes.* Morgan & Claypool Publishers.

[12] Margaret A Boden. 2004. *The creative mind: Myths and mechanisms.* Routledge.

[13] Olha Bondarenko and Ruud Janssen. 2005. Documents at hand: Learning from paper to improve digital technologies. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 121–130.

[14] Olha Bondarenko, Ruud Janssen, and Samuël Driessen. 2010. Requirements for the design of a personal document-management system. *Journal of the American Society for Information Science and Technology* 61, 3 (2010), 468–482.

[15] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2301–2309.

[16] John D Bransford and Barry S Stein. 1993. The IDEAL problem solver. (1993).

[17] Frederik Brudy, David Ledo, Michel Pahud, Nathalie Henry Riche, Christian Holz, Anand Waghmare, Hemant Bhaskar Surale, Marcus Peinado, Xiaokuan Zhang, Shannon Joyner, Badrish Chandramouli, Umar Farooq Minhas, Jonathan Goldstein, William Buxton, and Ken Hinckley. 2020. SurfaceFleet: Exploring Distributed Interactions Unbounded from Device, Application, User, and Time. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20).* Association for Computing Machinery, New York, NY, USA, 7–21. https://doi.org/10.1145/3379337.3415874

[18] Pamela Burnard, Anna Craft, Teresa Cremin, Bernadette Duffy, Ruth Hanson, Jean Keene, Lindsay Haynes, and Dawn Burns. 2006. Documenting 'possibility thinking': A journey of collaborative enquiry. *International journal of early years education* 14, 3 (2006), 243–262.

[19] Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design.* Morgan kaufmann.

[20] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis.* sage.

[21] Dar-Yen Chen, Hmrishav Bandyopadhyay, Kai Zou, and Yi-Zhe Song. 2025. Normalized Attention Guidance: Universal Negative Guidance for Diffusion Model. *arXiv preprint arXiv:2505.21179* (2025).

[22] DaEun Choi, Sumin Hong, Jeongeon Park, John Joon Young Chung, and Juho Kim. 2024. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 1055, 25 pages. https://doi.org/10.1145/3613904.3642794

[23] Stella Christie and Dedre Gentner. 2014. Language helps children succeed on a classic analogy task. *Cognitive science* 38, 2 (2014), 383–397.

[24] John Joon Young Chung and Eytan Adar. 2023. PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23).* Association for Computing Machinery, New York, NY, USA, Article 6, 17 pages. https://doi.org/10.1145/3586183.3606777

[25] ComfyUI. 2024. . https://www.comfy.org/

[26] Nigel Cross. 1997. Descriptive models of creative design: application to an example. *Design studies* 18, 4 (1997), 427–440.

[27] Nigel Cross. 2011. Understanding design thinking. *Notes on doctoral re-search in design: Contributions from the Politecnico Di Milano. FrancoAngeli* (2011), p19–37.

[28] 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.* 1–17.

[29] Alan Dix and Layda Gongora. 2011. Externalisation and design. In *Procedings of the second conference on creativity and innovation in design.* 31–42.

[30] Paul Dourish. 2001. *Where the action is: the foundations of embodied interaction.* MIT press.

[31] Claudia Eckert and Martin Stacey. 2000. Sources of inspiration: a language of design. *Design studies* 21, 5 (2000), 523–538.

[32] Figma. 2025. *Introducing Figma Weave: The next generation of AI-native creation at Figma.* https://www.figma.com/blog/welcome-weavy-to-figma/

[33] FLORA. 2025. *FLORA.* https://www.florafauna.ai/academy/quickstart

[34] Christiane Floyd. 1984. A systematic look at prototyping. In *Approaches to prototyping.* Springer, 1–18.

[35] FLUX.1. 2024. *FLUX.1.* https://getimg.ai/models/flux

[36] Jonas Frich, Michael Mose Biskjaer, Lindsay MacDonald Vermeulen, Christian Remy, and Peter Dalsgaard. 2019. Strategies in Creative Professionals' Use of Digital Tools Across Domains. In *Proceedings of the 2019 Conference on Creativity and Cognition.* 210–221.

[37] William Gaver, Peter Gall Krogh, Andy Boucher, and David Chatting. 2022. Emergence as a feature of practice-based design research. In *Proceedings of the 2022 ACM designing interactive systems conference.* 517–526.

[38] John S Gero. 1994. Computational models of creative design processes. In *Artificial intelligence and creativity: An interdisciplinary approach.* Springer, 269–281.

[39] Vinod Goel and Peter Pirolli. 1992. The structure of design problem spaces. *Cognitive science* 16, 3 (1992), 395–429.

[40] William JJ Gordon. 1961. Synectics: The development of creative capacity. (1961).

[41] Saul Greenberg, Sheelagh Carpendale, Nicolai Marquardt, and Bill Buxton. 2011. *Sketching User Experiences: The Workbook.* Elsevier.

[42] Sarah Henderson and Ananth Srinivasan. 2009. An empirical analysis of personal digital document structures. In *Symposium on Human Interface.* Springer, 394–403.

[43] 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* (San Jose, California, USA) *(CHI '07).* Association for Computing Machinery, New York, NY, USA, 251–260. https://doi.org/10.1145/1240624.1240666

[44] Jonathan Ho and Tim Salimans. 2022. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598* (2022).

[45] Ruey-Yun Horng, Ching-Wen Wang, Yung-Chieh Yen, Chia-Ying Lu, and Chien-Tao Li. 2021. A behavioural measure of imagination based on conceptual combination theory. *Creativity Research Journal* 33, 4 (2021), 376–387.

[46] Scott G Isaksen, K Brian Dorval, and Donald J Treffinger. 2010. *Creative approaches to problem solving: A framework for innovation and change.* SAGE publications.

[47] Alexander Ivanov, David Ledo, Tovi Grossman, George Fitzmaurice, and Fraser Anderson. 2022. MoodCubes: Immersive Spaces for Collecting, Discovering and Envisioning Inspiration Materials. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 189–203. https://doi.org/10.1145/3532106.3533565

[48] William Jones. 2010. *Keeping found things found: The study and practice of personal information management.* Morgan Kaufmann.

[49] James C Kaufman. 2016. *Creativity 101.* Springer publishing company.

[50] Tom Kelley. 2001. *The art of innovation: Lessons in creativity from IDEO, America's leading design firm.* Vol. 10. Crown Currency.

[51] Alison Kidd. 1994. The marks are on the knowledge worker. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 186–191.

[52] 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, Article 45 (may 2020), 27 pages. https://doi.org/10.1145/3392850

[53] Janin Koch, Nicolas Taffin, Andrés Lucero, and Wendy E. Mackay. 2020. SemanticCollage: Enriching Digital Mood Board Design with Semantic Labels. In *DIS '20 - Designing Interactive Systems Conference 2020*. ACM, 407. https://doi.org/10.1145/3357236.3395494

[54] Magda Kochanowska and Weronika Rochacka Gagliardi. 2021. The double diamond model: In pursuit of simplicity and flexibility. In *Perspectives on design II: Research, education and practice*. Springer, 19–32.

[55] Vladimir Kulikov, Matan Kleiner, Inbar Huberman-Spiegelglas, and Tomer Michaeli. 2024. FlowEdit: Inversion-Free Text-Based Editing Using Pre-Trained Flow Models. *arXiv preprint arXiv:2412.08629* (2024).

[56] David Ledo. 2025. Generative Rotoscoping: A First-Person Autobiographical Exploration on Generative Video-to-Video Practices. In *Proceedings of the 2025 Conference on Creativity and Cognition*. 931–948.

[57] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation strategies for HCI toolkit research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–17.

[58] Joanne Leong, David Ledo, Thomas Driscoll, Tovi Grossman, George Fitzmaurice, and Fraser Anderson. 2025. Paratrouper: Exploratory Creation of Character Cast Visuals Using Generative AI. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[59] Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction (TOCHI)* 15, 2 (2008), 1–27.

[60] Haichuan Lin, Yilin Ye, Jiazhi Xia, and Wei Zeng. 2025. SketchFlex: Facilitating Spatial-Semantic Coherence in Text-to-Image Generation with Region-Based Sketches. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 546, 19 pages. https://doi.org/10.1145/3706598.3713801

[61] Robert S Lockhart, Mary Lamon, and Mary L Gick. 1988. Conceptual transfer insimple insight problems. *Memory & Cognition* 16, 1 (1988), 36–44.

[62] Brian Logan and Tim Smithers. 2013. Creativity and design as exploration. In *Modeling creativity and knowledge-based creative design*. Psychology Press, 139–175.

[63] Yuwen Lu, Alan Leung, Amanda Swearngin, Jeffrey Nichols, and Titus Barik. 2025. Misty: UI Prototyping Through Interactive Conceptual Blending. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1108, 17 pages. https://doi.org/10.1145/3706598.3713924

[64] Andrés Lucero. 2012. Framing, Aligning, Paradoxing, Abstracting, and Directing: How Design Mood Boards Work. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. Association for Computing Machinery, New York, NY, USA, 438–447. https://doi.org/10.1145/2317956.2318021

[65] Andrés Lucero. 2012. Framing, aligning, paradoxing, abstracting, and directing: how design mood boards work. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) *(DIS '12)*. Association for Computing Machinery, New York, NY, USA, 438–447. https://doi.org/10.1145/2317956.2318021

[66] Matthew C Makel and Jonathan A Plucker. 2014. Creativity is more than novelty: Reconsidering replication as a creativity act. (2014).

[67] Thomas W Malone. 1983. How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Information Systems (TOIS)* 1, 1 (1983), 99–112.

[68] 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 *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–19.

[69] Ginés Hidalgo Martınez. 2019. *Openpose: Whole-body pose estimation.* Ph. D. Dissertation. Carnegie Mellon University Pittsburgh, PA, USA.

[70] Deana Mcdonagh and Ian Storer. 2004. Mood Boards as a Design Catalyst and Resource: Researching an Under-Researched Area. *The Design Journal* 7, 3 (2004), 16–31. https://doi.org/10.2752/146069204789338424 arXiv:https://doi.org/10.2752/146069204789338424

[71] Midjourney. 2024. *Midjourney Documentation.* https://docs.midjourney.com/

[72] Milanote. 2025. *Milanote.* https://milanote.com/

[73] Miro. 2025. *Miro.* https://miro.com/

[74] Michele I Mobley, Lesli M Doares, and Michael D Mumford. 1992. Process analytic models of creative capacities: Evidence for the combination and reorganization process. *Creativity Research Journal* 5, 2 (1992), 125–155.

[75] Mick Mahler (Mick Mumpitz). 2024. . https://www.youtube.com/@mickmumpitz

[76] Mural. 2025. *FLORA.* https://www.mural.co/use-case/design-canvas

[77] Brad Myers, Scott E Hudson, and Randy Pausch. 2000. Past, present, and future of user interface software tools. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 1 (2000), 3–28.

[78] Carman Neustaedter and Phoebe Sengers. 2012. Autobiographical design in HCI research: designing and learning through use-it-yourself. In *Proceedings of the Designing Interactive Systems Conference*. 514–523.

[79] Stellan Ohlsson. 1992. Information-processing explanations of insight and related phenomena. *Advances in the psychology of thinking* (1992), 1–44.

[80] Dan R Olsen Jr. 2007. Evaluating user interface systems research. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*. 251–258.

[81] OpenAI. 2024. *DALLE·3.* https://openai.com/index/dall-e-3/

[82] OpenAI. 2025. *ChatGPT.* https://chatgpt.com/overview/

[83] Eyal Oren. 2006. An Overview of Information Management and Knowledge Work Studies: Lessons for the Semantic Desktop. *SemDesk* (2006).

[84] Eyal Oren. 2006. An Overview of Information Management and Knowledge Work Studies: Lessons for the Semantic Desktop. *SemDesk* (2006). Publisher: Citeseer.

[85] Andrew Ortony. 1979. Beyond literal similarity. *Psychological review* 86, 3 (1979), 161.

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

[87] 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*. 804–818.

[88] Qualifications and Curriculum Authority. 2005. *Creativity: Find it, Promote It!—Promoting Pupils' Creative Thinking and Behaviour across the Curriculum at Key Stages 1, 2 and 3—Practical Materials for Schools.* Qualifications and Curriculum Authority, London.

[89] Albert Rothenberg. 1979. Einstein's creative thinking and the general theory of relativity: a documented report. *The American journal of psychiatry* 136, 1 (1979), 38–43.

[90] RunwayML. 2024. *Gen-3 Alpha.* https://runwayml.com/

[91] Vishnu Sarukkai, Lu Yuan, Mia Tang, Maneesh Agrawala, and Kayvon Fatahalian. 2024. Block and Detail: Scaffolding Sketch-to-Image Generation. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 33, 13 pages. https://doi.org/10.1145/3654777.3676444

[92] Keith Sawyer. 2013. *Zig zag: The surprising path to greater creativity.* John Wiley & Sons.

[93] R Keith Sawyer. 2021. The iterative and improvisational nature of the creative process. *Journal of Creativity* 31 (2021), 100002.

[94] Robert Keith Sawyer and Danah Henriksen. 2024. *Explaining creativity: The science of human innovation.* Oxford university press.

[95] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action.* Routledge.

[96] Ginamarie Scott, Lyle E Leritz, and Michael D Mumford. 2004. Types of creativity training: Approaches and their effectiveness. *The Journal of Creative Behavior* 38, 3 (2004), 149–179.

[97] Abigail J Sellen and Richard HR Harper. 2003. *The myth of the paperless office.* MIT press.

[98] Yulin Shen, Yifei Shen, Jiawen Cheng, Chutian Jiang, Mingming Fan, and Zeyu Wang. 2024. Neural Canvas: Supporting Scenic Design Prototyping by Integrating 3D Sketching and Generative AI *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1056, 18 pages. https:

//doi.org/10.1145/3613904.3642096

[99] Edward E Smith and Daniel N Osherson. 1984. Conceptual combination with prototype concepts. *Cognitive science* 8, 4 (1984), 337–361.

[100] Pieter Jan Stappers and Elisa Giaccardi. 2017. Research through design. In *The encyclopedia of human-computer interaction*. The Interaction Design Foundation, 1–94.

[101] Evgeny Stemasov, David Ledo, George Fitzmaurice, and Fraser Anderson. 2023. Immersive Sampling: Exploring Sampling for Future Creative Practices in Media-Rich, Immersive Spaces. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 212–229.

[102] Robert J Sternberg. 2003. The development of creativity as a decision-making process. *Creativity and development* 1 (2003), 91–138.

[103] Robert J Sternberg. 2020. Stalking the elusive creativity quark: Toward a comprehensive theory of creativity. In *New directions in aesthetics, creativity and the arts*. Routledge, 79–104.

[104] Luming Tang, Nataniel Ruiz, Qinghao Chu, Yuanzhen Li, Aleksander Holynski, David E. Jacobs, Bharath Hariharan, Yael Pritch, Neal Wadhwa, Kfir Aberman, and Michael Rubinstein. 2024. RealFill: Reference-Driven Generation for Authentic Image Completion. *ACM Trans. Graph.* 43, 4, Article 135 (July 2024), 12 pages. https://doi.org/10.1145/3658237

[105] tldraw. 2025. *tldraw*. https://tldraw.dev/quick-start

[106] W Edgar Vinacke. 1952. The psychology of thinking. (1952).

[107] Matteo Spinelli (Latent Vision). 2024. . https://www.youtube.com/@latentvision

[108] G Wallas. 1926. The art of thought. *Franklin Watts* (1926).

[109] Jagoda Walny. 2016. Thinking with Sketches: Leveraging Everyday Use of Visuals for Information Visualization. (2016).

[110] 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* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 136–146. https://doi.org/10.1145/3563657.3596072

[111] Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Feiwu Yu, Haiming Zhao, Jianxiao Yang, Jianyuan Zeng, Jiayu Wang, Jingfeng Zhang, Jingren Zhou, Jinkai Wang, Jixuan Chen, Kai Zhu, Kang Zhao, Keyu Yan, Lianghua Huang, Mengyang Feng, Ningyi Zhang, Pandeng Li, Pingyu Wu, Ruihang Chu, Ruili Feng, Shiwei Zhang, Siyang Sun, Tao Fang, Tianxing Wang, Tianyi Gui, Tingyu Weng, Tong Shen, Wei Lin, Wei Wang, Wei Wang, Wenmeng Zhou, Wente Wang, Wenting Shen, Wenyuan Yu, Xianzhong Shi, Xiaoming Huang, Xin Xu, Yan Kou, Yangyu Lv, Yifei Li, Yijing Liu, Yiming Wang, Yingya Zhang, Yitong Huang, Yong Li, You Wu, Yu Liu, Yulin Pan, Yun Zheng, Yuntao Hong, Yupeng Shi, Yutong Feng, Zeyinzi Jiang, Zhen Han, Zhi-Fan Wu, and Ziyu Liu. 2025. Wan: Open and Advanced Large-Scale Video Generative Models. arXiv:2503.20314 [cs.CV] https://arxiv.org/abs/2503.20314

[112] Wen-Fan Wang, Chien-Ting Lu, Nil Ponsa i Campanyà, Bing-Yu Chen, and Mike Y. Chen. 2025. AIdeation: Designing a Human-AI Collaborative Ideation System for Concept Designers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 21, 28 pages. https://doi.org/10.1145/3706598.3714148

[113] Robert Weisberg. 1986. *Creativity: Genius and other myths*. WH Freeman/Times Books/Henry Holt & Co.

[114] Robert W Weisberg. 1993. Creativity: Beyond the myth of genius. *(No Title)* (1993).

[115] Mikael Wiberg and Erik Stolterman. 2014. What makes a prototype novel? A knowledge contribution concern for interaction design research. In *Proceedings of the 8th Nordic conference on human-computer interaction: fun, fast, foundational*. 531–540.

[116] Shaojin Wu, Mengqi Huang, Yufeng Cheng, Wenxu Wu, Jiahe Tian, Yiming Luo, Fei Ding, and Qian He. 2025. USO: Unified Style and Subject-Driven Generation via Disentangled and Reward Learning. arXiv:2508.18966 [cs.CV] https://arxiv.org/abs/2508.18966

[117] Haijun Xia, Bruno Araujo, Tovi Grossman, and Daniel Wigdor. 2016. Object-oriented drawing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4610–4621.

[118] Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, and Lu Yuan. 2024. Florence-2: Advancing a unified representation for a variety of vision tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4818–4829.

[119] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. 2024. Depth anything v2. *Advances in Neural Information Processing Systems* 37 (2024), 21875–21911.

[120] Chengzhi Zhang, Weijie Wang, Paul Pangaro, Nikolas Martelaro, and Daragh Byrne. 2023. Generative Image AI Using Design Sketches as input: Opportunities and Challenges. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) *(C&C '23)*. Association for Computing Machinery, New York, NY, USA, 254–261. https://doi.org/10.1145/3591196.3596820

[121] Lei Zhang, Jin Pan, Jacob Gettig, Steve Oney, and Anhong Guo. 2024. VRCopilot: Authoring 3D Layouts with Generative AI Models in VR. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 96, 13 pages. https://doi.org/10.1145/3654777.3676451

[122] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3836–3847.

[123] Shihao Zhao, Dongdong Chen, Yen-Chun Chen, Jianmin Bao, Shaozhe Hao, Lu Yuan, and Kwan-Yee K Wong. 2024. Uni-controlnet: All-in-one control to text-to-image diffusion models. *Advances in Neural Information Processing Systems* 36 (2024).

[124] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 493–502.

	PARAMETER	WHAT IT SETS	ACTION	RATIONALE	OPERATION ON EASELS
GENERALIZED IMAGE/VIDEO WORKFLOWS	**Width**	*Image dimensions*	*Abstract*	*Depending on the model's training, there are ideal sizes (e.g., SDXL is 1024x1024). If the resolution is too low the model struggles to render details in place, if too high it can become slow or incoherent.*	*Fixed sizes based on aspect ratios: 1:1, 3:2, 4:3, 16:9, 2:3, 9:16, 3:4. Ability to set a starting image.*
	Height				
	Latent	*Starting point for inference*			
	Checkpoint	*Model of what can be rendered*	*Preset*	*This is most common with Stable Diffusion models. We pre-chose representative styles.*	*Style selection for SDXL.*
	Prompt	*Conditioning of what to render*	*Expose*	*Prompts define what can be drawn. Negative prompt mitigates unwanted features.*	*Textboxes for prompting. Keywords to trigger LoRAs.*
	Guidance Scale	*Adherence to the prompt*	*Abstract + Expose*	*Optimal values vary across models. The guidance scale can also be dependent on the number of steps.*	*Slider named "Prompt Adherence" that maps to model-specific ranges.*
	Seed	*Defines initial noise used for generation*	*Expose*	*Fixed seeds preserve more consistency versus changing the seed leads to variation.*	*Textbox that only accepts numbers.*
	Sampler/ Scheduler	*How denoising becomes an image*	*Hide*	*Best sampler/scheduler depends on the model or as the result of multiple experimentations.*	*N/A*
	Inference Steps	*Denoising steps to render an image*	*Hide*	*Models have a recommended number of inference steps.*	*Hidden.*
	Denoise	*How much to denoise the image*	*Abstract + Expose*	*'Denoise' is a complex concept, focus should be whether to render over an image or not.*	*A slider named "Preserve Original" for starting images.*
	Batch Size	*How many images to generate*	*Hide*	*Always generate a single result. Variations can be achieved with a new seed value.*	*N/A*
	Shift	*How sampler handles noise over time*	*Hide*	*This value is often preset by model.*	*N/A*
	Text Encoder	*Conversion from prompt to visual*	*Hide*	*Each model has a set of predefined encoders.*	*N/A*
	Variable Autoencoder	*Translation from pixel to latent space*	*Hide*	*Each model has a set of predefined VAEs.*	*N/A*
LORAS	**Style LoRA Checkpoints**	*The kinds of images and styles that can be created based trained datasets*	*Preset + Expose*	*We curate style LoRAs that can be mixed (e.g., realism, 3D). Strength should be adjustable. Editing models one can use LoRAs for better results (e.g., relighting), and video models can use LoRAs for camera movement.*	***Draw:** on-off* ***Paint/Trace:** Radial sliders* ***Modify:** Tabs for single edits* ***Animate:** Camera motions.*
	Style LoRA Strength				
	Efficiency LoRAs	*Reduction of number of inference steps.*	*Hide*	*Some LoRAs reduce total steps. Faster renders lead to less motion in videos.*	***Animate:** expose 'quality' (low, medium, high).*
DETAILS	**Dishonesty Factor**	*How much noise to inject to add details.*	*Abstract + Expose*	*The dishonesty factor (-0.05 to 0) can increase in details without damaging the result.*	*A slider renamed "Add Details"*
	Start/End	*Inference steps to apply noise injection*	*Hide*	*Adds complexity, not directly noticeable.*	*N/A*
CONTROLNET	**Preprocessed Image**	*Structural composition to work with*	*Abstract + Expose*	*We choose a set of representative preprocessors to compute when media is imported.*	*Control images selection within the structure reference.*
	Checkpoint	*How to process the image*	*Hide*	*ControlNet Union supports many types of preprocessed images.*	*N/A*
	Start/End	*Inference steps that apply ControlNet*	*Fixed*	*ControlNets work best when starting at 0. End values can be fixed.*	*N/A*
	Strength	*The strength of the ControlNet*	*Abstract*	*The strength of ControlNet determines how much to follow the structure.*	*A slider for the strength*
IMAGE REFERENCES	**Image**	*Reference Image*	*Expose*	*The image to use as reference.*	*A box to add an image*
	Checkpoint	*Style model*	*Hide*	*We use Flux REDUX / IP-Adapter Plus (SDXL).*	*N/A*
	Clip Vision	*Vision model*	*Hide*	*Each model has a compatible CLIP Vision*	*N/A*
	Attention Mask	*Determines the area for reference to affect*	*Abstract + Expose*	*SDXL only with IP-Adapter, black and white images define areas of interest.*	*A drawing tool to set the masks over the start image*
	Strength	*Reference strength*	*Abstract*	*One can set how much effect a reference has*	*A slider for the strength*
	Type	*How references act*	*Hide*	*We preset IP-Adapter to "strong style transfer"*	*N/A*
	Start/End	*Inference steps that apply reference*	*Hide*	*We apply references to the full inference.*	*N/A*
VID	**Start Frame**	*First video frame*	*Expose*	*Optional start frame*	*A box to add an image*
	End Frame	*Last video frame*	*Expose*	*Optional end frame*	*A box to add an image*

Table 3: Rationale of workflow parameters as they get converted into easels.

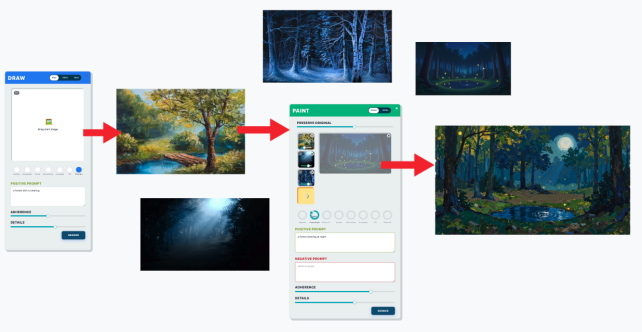

Figure 13: Designing the Setting. A forest is generated and then used as a style reference along other images of forests.

A Usage Scenario

To illustrate how some of the features of Atelier come together, we present a use case scenario for Atelier. Fiona is a filmmaker who wants to ideate a short fantasy film concept about a warrior that kills a dragon.

Fiona opens Atelier and import images of dragons and lizards, epic hero shots, and fantasy scenes that carry the aesthetic of world they want to create (Figure 11). Looking at these images arranged in the canvas, acting as a mood board, they get a sense of the atmosphere of the story.

Figure 11: Initial canvas with drag-and-dropped inspiration images.

Fiona wants to settle on a protagonist first. She writes out a few descriptions and uses **Quick Sketch** to rapidly render images and likes a concept for the prompt "female warrior wearing a cape" but dislikes how the character is facing directly forward. She uses **Remove Background**, then **Sculpt** to turn her into a rough 3D model and is able to adjust her warrior to any angle and get still captures of different views.

Figure 12: The 3D model is used to create different angled views of warrior.

She wants to see her warrior in a scene, so she opens a **Collage** easel and arranges an image of her warrior on a scene of a lake from her inspiration images for a close up shot. Taking this glued image, she then opens a **Trace** easel, and describes her original image as "a woman wearing a red cape in front of a lake" and her target image "a woman wearing a red cape, painting style." The generated image does not have the background she wanted, but Fiona likes the newly stylized version of the warrior. She uses **Remove Background** again and then saves the image into a new collection titled "Warrior."

She now focuses on the setting for the scene: a forest. She continues building on painting styles, using the **Draw** easel with "a forest with a clearing" (Figure 13). She likes the result, but wants it to be night time, so she uses **Revision**. One of the results surprises her, giving her new ideas, and she is drawn to the spooky feel and decides to move in that direction instead. She opens a **Paint** easel and references the spooky forest and other images with the prompt "a forest clearing at night". She likes the paint style of the first generation and includes it as a style for subsequent generations. She opens the **Collage** easel and composes another close up of the

Figure 14: Exploring camera angles and staging of the warrior character

warrior image taken from the collection and the forest scene. Fiona is unsure if this combination works well, so she uses **Revision** to try out different views with directions (Figure 14) such as "make the woman face forward" and "show woman from waist up". From these, she likes some of the zoomed out views, so Fiona uses **Remove Background** again to extract the warrior in the angles she wants. Now she has a few angles and zoom levels for her scenes. Next is the dragon. Going back to the earlier imported images of dragons and lizards, Fiona plays around with different easels and combinations to generate dragon designs (Figure 15). She eventually lands on a colour scheme that is gray and dark. This inspires her to think of

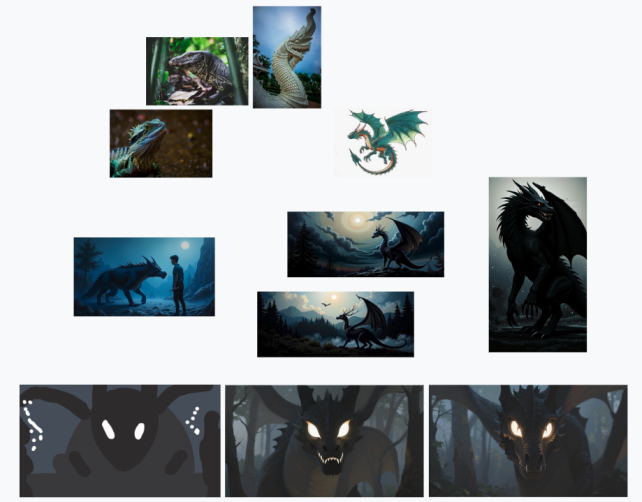

Figure 15: Character generation explorations for Dragon.

the dragon scene as one that is thematically darker than the rest. She uses these gray colors in the **Sketch** easel to draw out her dragon, and then **Trace**, this time with her drawing as the input. After iterating on the parameters, she gets a design she is happy with. She then uses the **Modify** easel to get 16:9 aspect ratio shots with a variety of different camera angles and lighting. Her favorite ones get added to a new "Dragon" collection.

Observing the canvas, Fiona has an idea for a sequence: the warrior standing in the forest, the camera panning from her face around to her back, then pushing into the trees until the dragon emerges. She chooses the images of the warrior in the forest, the empty forest, the dragon approaching, and finally a close-up of the dragon, and adds them to the gallery. Reviewing the sequence, she realizes she is missing a crucial frame that positions the camera behind the warrior, and uses **Revision** to create it. From these frames (Figure 16), she can use the **Animate** easel to pass in these images as start and end images to create a continuous video.

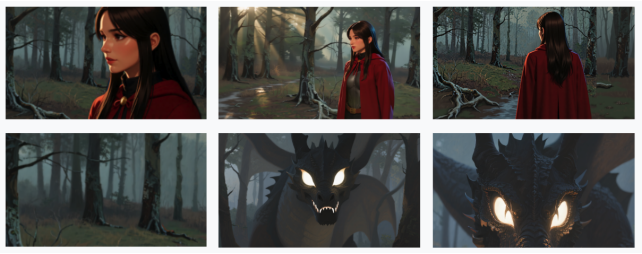

Figure 16: Final frames in the storyboard showing a female warrior in the forest as a dragon emerges.

B Implementation Details

B.1 Small Workflows and Premade Workflows

Many of the quick operations are done with ComfyUI workflows with a few nodes or reusing nodes available online, such as using template remove background, extract element (GroundingDINO[10]), palette, upscale, extend (outpaint) and stencil (control images). Several of these actions are pared down versions of existing easels with no extra settings (Quick Sketch and Draw; Revision, View and Modify, Quick Animate and Animate). Sculpt employs the Hunyuan3D wrapper workflow [11] to go from image to 3D.

B.2 Precomputing Meta-data

Whenever an image is uploaded, we run pre-processing steps to store meta-data and avoid recalculations. Images are captioned with Florence2 [118], and images are preprocessed for ControlNets: OpenPose [69], DepthAnything V2 [119], Scribble, and Lineart. This metadata is later used for other functions such as search or supporting structure images in some easels.

B.3 Draw and Paint Workflows

The Draw easel uses the same workflows for Stable Diffusion XL and FLUX except features for reference images, negative prompt, etc. are not exposed. However, the Draw easel also supports Wan 2.2 as an image generator with our custom workflow.

B.3.1 Text-to-Image in the Wan 2.2 Model. The Wan 2.2 text-to-video model [111] uses a *'mixture of experts'* to create visuals. When creating videos, The *'high noise'* checkpoint is responsible for the movement, and the *'low noise'* model creates intelligible high-quality visuals from it. To create an image, we use the *'low noise'* model for a single frame. The Wan model for image generation can render realistic poses and expressions, which is harder to achieve in traditional image models.

We use the Wan Lightning LoRA which is originally designed for reducing the number of inference steps from 20 to 4 at a classifier-free-guidance value range of 1.0 to 1.5. We found that using the Lightning LoRA at 20 steps leads to much higher quality renders. We added an additional layer of Normalized Attention Guidance at a strength of 11, to enable negative prompts and Lying Sigmas for details. We found that the Lying Sigma value behaves differently in the video model compared to typical image models as it seems to also affect the stylization.

B.3.2 FLUX Generations with References, Style, and Structure. Paint works by loading the FLUX model and encoders, then embedding the active LoRAs. We tie three reference images to FLUX Redux[12] using the AdvancedRefluxControl node[13] which adds further controllability. The AdvancedRefluxControl does not have an inactive state, so to avoid switches or automatic rewiring, we modified the code for the node and added a condition to return 0. To control the structure we expose one of the control images from the pre-processed metadata at fixed end percentages as ControlNet using ControlNetUnion [123] (pose: 0.9, depth 0.7, and lineart 0.4). We also use Normalized Attention Guidance for negative prompt and Lying Sigmas for details. The generation is done with 8 inference steps leveraging FLUX Turbo[5].

[10] https://github.com/storyicon/comfyui_segment_anything
[11] https://github.com/kijai/ComfyUI-Hunyuan3DWrapper
[12] https://huggingface.co/black-forest-labs/FLUX.1-Redux-dev
[13] https://github.com/kaibioinfo/ComfyUI_AdvancedRefluxControl

Because FLUX Redux impacts the conditioning, image references affect both style and composition which can conflict with ControlNet for structure. We created an alternative version of the workflow where one of the image references ties to ByteDance USO [116] to define only style while preserving the compatibility with Redux and with ControlNets.

B.4 Modify Workflow

Modify uses Flux Kontext to edit existing images, such as changing cinematographic attributes. We add a set of LoRAs for changing camera angle, relighting, and styles. Unique to this workflow is the ability to recompose an image into a completely new aspect ratio.

B.5 Trace Workflows

We use two workflows to support Trace - one for FLUX and one for Wan 2.2, both based on workflows and custom nodes by Github user Log(td)[14] with a technique called FlowEdit [55]. FlowEdit reverse-engineers a tagged starting image and reconstructs it based on a new prompt. This approach came about before editing models such as FLUX Kontext and Qwen Edit, and thus can transform images in interesting ways. The FLUX workflow is based on Log(td)'s Fluxtapoz node package[15]. We modified the sample workflow by adding our set of custom LoRAs, ControlNet, and USO style images [116]. In our implementation we use 20 inference steps.

For FlowEdit in Wan 2.2 image generation, we created a workflow combining our text-to-image workflow with a workflow inspired by Zack Abrams[16], who modified a Hunyuan Video FlowEdit workflow[17] for the Wan 2.1 image-to-video model.

[14] https://github.com/logtd
[15] https://github.com/logtd/ComfyUI-Fluxtapoz
[16] https://tinyurl.com/AbramsWanFlowEditWorkflow
[17] https://github.com/logtd/ComfyUI-HunyuanLoom

C Questionnaire

To understand how participants perceived the role and qualities of generative systems within their creative workflows, we asked them to rate a set of statements on a 5-point Likert scale (Strongly Disagree → Strongly Agree). Below we provide the full wording of each questionnaire item alongside the label used in the paper.

(1) **Ownership**: I feel ownership of what I create with ____.
(2) **Agency**: I feel control and agency using ____.
(3) **Early Stage Thinking**: ____ supports my early stage thinking.
(4) **Late Stage Thinking**: ____ supports my late stage thinking.
(5) **Inspiration**: I use ____ for inspiration.
(6) **Final Product**: Creations from ____ could show up in a final product.
(7) **Malleability**: ____ feels malleable.
(8) **Integration to Process**: ____ integrates smoothly into my creative process.
(9) **Integration to Tools**: ____ integrates smoothly with my other tools.
(10) **End-to-End Support**: ____ supports end-to-end workflows.
(11) **Reflects Intent**: I feel confident that ____ will reflect my design intent.
(12) **Enjoyment**: I enjoy using ____.
(13) **Fun**: ____ feels fun to use.
(14) **Professional-feeling**: ____ feels professional grade.
(15) **Ability to Fix Outputs**: When I get an undesired output with ____, I feel I can easily fix it to get what I want.
(16) **Expressiveness**: ____ feels expressive.
(17) **Seen as Tool**: I see ____ as tools.
(18) **Seen as Collaborator**: I see ____ as collaborators.
(19) **Empowerment**: ____ makes me feel empowered.
(20) **Creativity**: ____ makes me feel creative.