Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Design and Evaluation of a Stylistically Personalized Private Generative AI Workflow

Master's Thesis

2026
Bc. Ján Hučko

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Design and Evaluation of a Stylistically Personalized Private Generative AI Workflow

Master's Thesis

Study Programme:	Cognitive Science
Field of Study:	Computer Science
Department:	Department of Applied Informatics
Supervisor:	doc. RNDr. Zuzana Berger Haladova, PhD

Bratislava, 2026
Bc. Ján Hučko

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Ján Hučko
Študijný program: kognitívna veda (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Design and Evaluation of a Stylistically Personalized Private Generative AI Workflow
Návrh a vyhodnotenie štylisticky personalizovaného súkromného generatívneho AI pracovného postupu

Anotácia: Diplomová práca skúma vývoj súkromného generatívneho pracovného postupu s AI, ktorý je navrhnutý na podporu tvorivej autonómie vizuálnych umelcov. Systém je určený na učenie sa jedinečného štýlu umelca z malého portfólia diel a na integráciu vizuálnych aj textových podnetov. Bude tiež obsahovať nástroje, ktoré umelcom umožnia spresniť a prispôsobiť vygenerované výsledky.

Cieľ: Výskumná otázka znie: Ako môže súkromný, štylisticky personalizovaný generatívny pracovný postup s AI podporiť kreatívnu autonómiu a profesionálnu užitočnosť vizuálnych umelcov? Navrhované riešenie má 4 ciele: (1) Použiteľnosť profesionálmi, ktorí nie sú expertmi na AI. (2) Užitočnosť pri podpore bežných kreatívnych úloh, a nie pri produkcii finálneho obrazu. (3) Personalizácia – generovanie výstupov, ktoré autenticky odrážajú jedinečný štylistický podpis používateľa. (4). Kontrola a dôvera používateľa z pohľadu súkromia jeho údajov .

Literatúra:
Jiang, H. H., et al. (2023). AI Art and its Impact on Artists. AAAI/ACM Conference on AI, Ethics, and Society, 363–374.
Johnston, H., & Thue, D. (2024). Understanding Visual Artists' Values and Attitudes towards Collaboration, Technology, and AI. Graphics Interface, 1–9.
Khan, M. A. et al. (2025). Personalizing AI Art Boosts Credit, Not Beauty. PsyArXiv.

Vedúci: doc. RNDr. Zuzana Berger Haladová, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. RNDr. Tatiana Jajcayová, PhD.

Dátum zadania: 13.02.2026

Dátum schválenia: 15.02.2026

prof. Ing. Igor Farkaš, Dr.
garant študijného programu

.....
študent

.....
vedúci práce

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Bc. Ján Hučko
Study programme:	Cognitive Science (Single degree study, master II. deg., full time form)
Field of Study:	Computer Science
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Design and Evaluation of a Stylistically Personalized Private Generative AI Workflow

Annotation: The thesis investigates the development of a private generative AI workflow designed to support creative autonomy of visual artists. The system is intended to learn artist's distinctive style from a small portfolio of works and to integrate both visual and textual prompting. It will also include tools for iterative editing and re-prompting, enabling artists to refine and adapt generated results.

Aim: The research question is: How can a private, stylistically personalized generative AI workflow support the creative autonomy and professional utility of visual artists? The proposed solution has 4 objectives: (1) Usability - operable by professionals who are not AI experts. (2) Utility – demonstrate practical utility in supporting common creative tasks rather than final image production. (3) Personalization – capable of generating outputs that authentically reflect the unique stylistic signature of the individual user. (4). Control and Trust – must operate in a manner giving the user a sense of control and confidence in the privacy of their data.

Literature:
Jiang, H. H., et al. (2023). AI Art and its Impact on Artists. AAAI/ACM Conference on AI, Ethics, and Society, 363–374.
Johnston, H., & Thue, D. (2024). Understanding Visual Artists' Values and Attitudes towards Collaboration, Technology, and AI. Graphics Interface, 1–9.
Khan, M. A. et al. (2025). Personalizing AI Art Boosts Credit, Not Beauty. PsyArXiv.

Supervisor:	doc. RNDr. Zuzana Berger Haladová, PhD.
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	doc. RNDr. Tatiana Jajcayová, PhD.
Assigned:	13.02.2026
Approved:	15.02.2026

prof. Ing. Igor Farkaš, Dr.
Guarantor of Study Programme

.....
Student

.....
Supervisor

Affidavit: I declare that the diploma thesis on the topic "Design and Evaluation of a Stylistically Personalized Private Generative AI Workflow", including all its appendices and images, I wrote independently, using the literature listed in the attached references and artificial intelligence tools. I declare that I have used artificial intelligence tools in accordance with the relevant legal regulations, academic rights and freedoms, ethical and moral principles while maintaining academic integrity and that their use is appropriately marked in my thesis. In preparing this thesis, I used Codex, Claude Code, Gemini CLI, and locally run large language models for preparing supporting materials and the structure of the text, searching and summarizing sources, organizing transcript data, linguistic and translation support, and technical assistance. The author is responsible for the correctness of the final text.

Acknowledgments: I would like to thank my supervisor, doc. RNDr. Zuzana Berger Haladova, PhD, for her assistance and advice throughout this work. I am also grateful to Asst. Prof. Univ.Lektor Dr. Soheil Human, MSc, BSc, and the Sustainable Computing Lab at WU Vienna for their encouragement and for supporting the development of this project. My thanks go to RNDr. Kristína Malinovská, PhD., for her support since the earliest idea for this project, and to the participants for their interest, time, and trust. I am also deeply thankful to my friends and family for their support. Finally, I would like to thank my partner for being a steady positive force in many ways throughout this process.

Abstract

This thesis examines how a private, stylistically personalized generative AI workflow can support the creative autonomy and professional utility of visual artists. Instead of defining usefulness only by output quality, it examines how artists responded to generated and edited images and where those images fit, or failed to fit, within their creative processes. A local ComfyUI-based workflow was designed for private, stylistically personalized generation, evaluated with two artists in researcher-supported exploratory sessions, and then redesigned to add prompt mediation and prompt-based editing. The redesigned workflow was evaluated with one repeat participant and one additional participant.

The study uses a small exploratory qualitative design. Data came from screen and audio recordings, generated and edited images, debrief interviews, and post-session questionnaires. The workflow was designed and evaluated around questions of usability, creative utility, personalization and stylistic fidelity, and the combined area of control, trust, privacy, and acceptance.

Across the evaluations, the workflow was most useful when it supported continuation rather than replacement. Participants valued outputs as references, inspiration, source material, and starting points for further work, rather than as substitutes for their own creative practice. Personalization made generated images more recognizable and easier to relate to, and in some cases affected participants' sense of ownership over the outputs. Local deployment shaped what kinds of portfolio, client, or personally meaningful material participants felt able to use. Prompt mediation reduced some of the burden of writing model-effective prompts, while editing gave participants another way to continue from promising images.

The findings suggest that, in this kind of workflow, usefulness depends not only on isolated image quality but on whether outputs support a next step in creative practice. Within the limits of short, researcher-supported sessions and a small purposive sample, the thesis contributes a design and evaluation account of how privacy, personalization, prompt support, and editing can work together to make generated material more relevant, usable, and acceptable for artists.

Keywords: generative AI, stylistic personalization, human-AI co-creativity, diffusion models, LoRA, visual artists, data privacy, local deployment

Abstrakt

Táto diplomová práca skúma, ako môže súkromný, štýlovo personalizovaný workflow generatívnej umelej inteligencie podporiť tvorivú autonómiu a profesionálnu využiteľnosť vizuálnych umelcov. Namiesto toho, aby užitočnosť definovala výlučne na základe kvality výstupov, skúma, ako umelci reagovali na generované a upravené obrazy a kde tieto obrazy zapadali, alebo nezapadali, do ich tvorivých procesov. V rámci práce bol navrhnutý lokálny workflow založený na ComfyUI pre súkromné, štýlovo personalizované generovanie, ktorý bol vyhodnotený s dvoma umelcami v prieskumných sedeniach podporovaných výskumníkom a následne prepracovaný s cieľom pridať podporu pri tvorbe promptov a úpravy pomocou promptov. Prepracovaný workflow bol vyhodnotený s jedným opakovaným účastníkom a jedným ďalším účastníkom.

Štúdia využíva malý prieskumný kvalitatívny dizajn. Dáta pochádzali z obrazových a zvukových záznamov, generovaných a upravených obrazov, záverečných rozhovorov a dotazníkov vyplnených po sedeniach. Workflow bol navrhnutý a vyhodnotený z hľadiska použiteľnosti, tvorivej užitočnosti, personalizácie a štýlovej vernosti a spoločnej oblasti kontroly, dôvery, súkromia a akceptácie.

Naprieč hodnoteniami bol workflow najužitočnejší vtedy, keď podporoval pokračovanie tvorivého procesu, a nie jeho nahradenie. Účastníci vnímali výstupy skôr ako referencie, inšpiráciu, zdrojový materiál a východiská pre ďalšiu prácu než ako náhradu vlastnej tvorivej praxe. Personalizácia robila generované obrazy rozpoznateľnejšími a bližšími vlastnej práci a v niektorých prípadoch ovplyvnila aj pocit vlastníctva výstupov. Prevádzka v lokálnom prostredí ovplyvnila, aké portfóliové, klientské alebo osobne významné materiály boli účastníci ochotní použiť. Podpora pri tvorbe promptov znížila časť záťaže spojenej s písaním efektívnych promptov pre model, zatiaľ čo úpravy poskytli účastníkom ďalší spôsob, ako pokračovať od sľubných obrazov.

Zistenia naznačujú, že v tomto type workflowu užitočnosť nezávisí iba od izolovanej kvality obrazov, ale aj od toho, či výstupy podporujú ďalší krok v tvorivej praxi. V rámci obmedzení krátkych sedení podporovaných výskumníkom a malej účelovej vzorky práca prispieva opisom návrhu a hodnotenia toho, ako môžu súkromie, personalizácia, podpora pri tvorbe promptov a úpravy spoločne prispieť k tomu, aby bol generovaný materiál pre umelcov relevantnejší, použiteľnejší a prijateľnejší.

Kľúčové slová: generatívna umelá inteligencia, štýlová personalizácia, spolutvorba človeka a umelej inteligencie, difúzne modely, LoRA, vizuálni umelci, súkromie dát, prevádzka v lokálnom prostredí

Contents

List of Figures

List of Tables

Introduction

Generative image systems can now produce images that are visually detailed, varied, and quick to inspect. For artists, however, the usefulness of such systems is not decided only by image quality. A generated image may be useful as a reference, a source of inspiration, a way to test a visual direction, or material for later transformation. It may also be unhelpful if it is difficult to steer, if it does not fit the artist's process, if it requires sensitive material to be uploaded to an external service, or if it produces finished-looking images that sit awkwardly with the artist's own sense of labor and authorship.

This thesis starts from that practical tension. The problem is not whether generative AI can produce convincing images in general. The more specific question is how generated images can remain useful inside an artist's own creative process. This makes usefulness a workflow question rather than only an output-quality question. A system may need to be usable enough for exploratory work, while still giving the artist ways to understand and adjust what is happening. It may need to produce images that are relevant to the artist's own material, without treating personalization as a complete model of artistic identity. It also raises questions about how prompts, input images, trained style models, and generated outputs should be handled, especially when those materials are connected to an artist's portfolio, client work, or ongoing creative projects.

For this reason, the thesis treats generative AI as a workflow rather than only as a model and its outputs. The workflow developed here brings together local deployment, stylistic personalization, image generation, prompt support, and image transformation. These components are not treated as separate technical features only. They are evaluated in relation to whether the material produced by the system can support the artist's own judgment and later work.

The aim of the present research is to examine how a private, stylistically personalized generative AI workflow can support artists' creative work. The thesis is guided by the following research question:

> *How can a private, stylistically personalized generative AI workflow support the creative autonomy and professional utility of visual artists?*

The phrase "creative autonomy" is used here in a practical sense. It refers to whether

participants could use the workflow while still feeling that they were directing the creative process, rather than having the system determine it for them. The phrase "professional utility" does not mean that the system was evaluated as a replacement for professional production. It refers to whether the workflow produced material that could plausibly support creative work.

The question is approached through design, evaluation, redesign, and re-evaluation. A first local workflow was designed and evaluated with two artists in researcher-supported sessions. The findings from that evaluation then guided a redesign, which was evaluated in a second round with one repeat participant and one additional participant. In this redesign, personalized generation was kept, but the workflow was extended with prompt support and an editing path so that generated and existing images could be used as material for further work.

The research does not aim to show that the workflow is ready for broad independent adoption. Participants used prepared workflows in short supported sessions. They did not install the software, maintain the local environment, create their own personalized model components, or use the system across a long project. The study is instead concerned with how participants used the two workflows, where the workflows supported creative work, where they remained difficult to control, and what these limits suggest for designing generative AI systems for visual artists.

This scope is important because the kind of workflow studied here involves more than ordinary use of a web-based image generator. It depends on technical setup, personalization, interaction design, and data handling decisions that shape how the workflow can be used. The thesis therefore evaluates research prototypes rather than a packaged application that artists could adopt independently.

The thesis draws on artificial intelligence and computer science through the use of generative image models, personalization methods, local inference, and prompt mediation. It also draws on human-computer interaction by asking how artists understand, control, and incorporate the system into creative activity. The questions about creative use connect to the psychology of creative cognition, especially ideation, fixation, and mental models. Finally, the questions of privacy, ownership, agency, and authorship connect the technical artifact to philosophical and ethical concerns about human contribution in AI-assisted work. The thesis therefore treats the workflow as both a computational system and a tool used within situated creative practice.

Chapter 1

Technical Background

This chapter introduces the technical concepts needed to understand the workflow evaluated in the thesis. It is not meant to be a complete technical survey of generative image modeling. Instead, it focuses on the parts of the literature that matter for the design problem: image synthesis and editing, personalization and stylistic adaptation, interaction and control, prompt mediation, and local deployment.

1.1 Generative AI for Image Synthesis

Generative image synthesis refers to computational systems that produce new visual content from learned patterns in data. Deep learning approaches to image synthesis have included variational autoencoders, generative adversarial networks, autoregressive models, transformer-based systems, and diffusion models, each with different tradeoffs between image quality, likelihood modeling, controllability, and computational cost (Kaur et al., 2026; Chang et al., 2026). In text-to-image systems, generation is conditional: the model has to produce a coherent image while following the user's prompt or another input signal, including intended content, style, composition, or task constraints (Kaur et al., 2026; Cao et al., 2026; Fang, 2024).

Ho et al. (2020) describe diffusion models as learning to reverse a process that gradually adds noise to data. During training, clean images are progressively corrupted with noise, and the model learns to predict or remove that noise at different timesteps. During generation, the learned reverse process starts from random noise and moves iteratively toward image structure. They also showed that this approach could produce high-quality image samples and formalized the training objective used to learn the denoising process. Later work improved the efficiency and sample quality of denoising diffusion probabilistic models, while also clarifying the importance of the noise schedule, objective design, and sampling procedure (Nichol and Dhariwal, 2021; Chang et al., 2026). During generation, the reverse process can be conditioned on signals such as

text, reference images, masks, structure maps, or other guidance inputs (Cao et al., 2026; Huang et al., 2025).

Latent diffusion models made high-resolution diffusion-based generation more practical by performing denoising in a compressed latent representation rather than directly in pixel space, then decoding the result back into an image (Rombach et al., 2021). This approach helps explain why diffusion models became usable in creative software contexts rather than only in large-scale research settings, and it underlies many widely used open image-generation systems (Rombach et al., 2021; Podell et al., 2023). More recent model families build on related continuous-time and flow-matching formulations, where generation is still understood as a learned transformation from noise or latent states toward coherent images, but with different sampling trajectories and efficiency tradeoffs (Fan et al., 2025; Black Forest Labs et al., 2025).

Text-to-image generation is only one part of the current image-synthesis space. In many applied workflows, the goal is not a fully new image from scratch, but a transformation of an existing visual artifact. Diffusion-based image editing addresses this by conditioning generation on an input image as well as on an instruction, prompt, mask, reference, or structural guide (Huang et al., 2025; Yu, 2025). The literature describes a progression from latent-space manipulation and target-prompt editing toward instruction-guided and interactive editing, where the user can describe a desired change rather than only describe the desired final image (Yu, 2025; Huang et al., 2025). Editing can therefore be understood as constrained generation: the model is still synthesizing image content, but the synthesis is constrained by an existing image context and by the user's requested transformation (Huang et al., 2025; Yu, 2025). A repeated technical difficulty in editing is the balance between edit strength and preservation: stronger transformations may better satisfy the new instruction, but they can also cause unwanted drift in the source image's structure, identity, or unedited regions (Huang et al., 2025; Yu, 2025).

The distinction between generation and editing matters because different forms of conditioning support different kinds of control. Inpainting constrains the change to a selected region; image-to-image generation constrains the output by an input image; reference-based methods constrain identity, style, or subject matter; and structural-control methods constrain geometry through edges, depth, pose, segmentation, or similar representations (Huang et al., 2025; Zhang et al., 2023; Cao et al., 2026). In the wider literature, ControlNet is a representative example of this shift toward spatial conditioning: it adds trainable conditioning pathways to a pretrained diffusion model so that prompts can be combined with inputs such as sketches, depth maps, human poses, and segmentation maps (Zhang et al., 2023). Image synthesis is therefore no longer limited to prompt-based creation of new images; it increasingly includes controlled visual transformation (Huang et al., 2025; Yu, 2025; Cao et al., 2026).

Recent unified models further blur the boundary between generation and editing. For example, the FLUX.1 Kontext technical report describes it as a flow-matching model that conditions on both text and image context within a single architecture, so that text-to-image generation can be treated as the case where no context image is provided, while image editing uses an existing image as additional context (Black Forest Labs et al., 2025). Official FLUX.2 documentation describes the model family in similar practical terms, presenting it as suitable for image generation and editing with multi-reference support, and positioning FLUX.2 [klein] as an open-weight, fast model family intended for local or high-volume use (Black Forest Labs, 2026a,c). Since FLUX.2 [klein] does not currently have a peer-reviewed technical paper, model-specific claims about it are limited to official documentation.

1.2 Personalization and Stylistic Fine-Tuning

Personalization describes the adaptation of a generative model to a user-specified concept, such as a subject, identity, object, character, or visual style, rather than only to a generic text prompt (Zhang et al., 2025; Wei et al., 2025). In diffusion models, the concept is usually provided through one or more reference images and then encoded or learned as a conditioning signal, such as a new token, an image feature, or a parameter update (Zhang et al., 2025; Wei et al., 2025). A prompt can describe a general category, but it does not by itself create a persistent technical representation of a specific user's material, recurring motifs, or preferred rendering tendencies (Zhang et al., 2025; Wei et al., 2025). In the technical literature, "style" usually refers to visible regularities such as colors, textures, brush strokes, materials, lighting, and related aesthetic features that can be inferred from reference images (Zhang et al., 2025; Wei et al., 2025). It should not be read as a claim that the model captures an artist's full style in the richer human sense of intention, judgment, process, subject matter, and artistic development.

A central difficulty in personalization is balancing fidelity with alignment. A personalization method should preserve the distinctive features of the reference subject or style while still allowing the generated image to respond flexibly to new prompts and contexts (Zhang et al., 2025; Wei et al., 2025). If personalization is too strong, the system may overfit to the reference set and reproduce the original pose, background, composition, or other irrelevant details. If it is too weak, the generated image may remain plausible but lose the features that made the concept recognizable (Zhang et al., 2025; Wei et al., 2025).

Personalization methods can be grouped into methods that adapt the model for each new concept and methods that use a pre-trained reference pathway (Zhang et al., 2025; Wei et al., 2025). The first group is usually described as test-time fine-tuning: when a

new reference concept is introduced, the system optimizes a token, model component, or added module from one or more example images, often with a unique identifier that can be used in later prompts (Zhang et al., 2025; Cao et al., 2026). Textual Inversion is a compact example of this approach, learning a new token embedding while keeping the main text-to-image model frozen (Gal et al., 2022; Wei et al., 2025). DreamBooth instead fine-tunes model parameters so that a rare identifier can refer to a specific subject, while using regularization to reduce overfitting and preserve the model's broader class knowledge (Ruiz et al., 2022; Zhang et al., 2025). The second group, pre-trained adaptation, relies on encoders or adapters trained in advance so that a new reference image can be converted into conditioning information at inference time (Zhang et al., 2025; Wei et al., 2025). IP-Adapter is a representative example: it uses a pretrained image encoder and additional attention pathways to guide generation from reference-image features alongside the text prompt, without optimizing a new concept-specific model for each use case (Ye et al., 2023; Wei et al., 2025).

Low-Rank Adaptation, or LoRA, belongs to the fine-tuning side of this distinction, but it makes adaptation much more compact than full model fine-tuning. LoRA freezes most pretrained model weights and learns small low-rank update matrices instead of retraining the full model (Hu et al., 2021). In image-generation practice, this means that a base diffusion model can be adapted through a comparatively small subject or style module rather than through a separate full model checkpoint, as in heavier fine-tuning approaches such as DreamBooth (Ruiz et al., 2022; Hu et al., 2021; Zhang et al., 2025; Liu et al., 2024). This reduces the storage and training burden of personalization, which is especially relevant when multiple artist-specific styles or subjects need to be kept available in the same workflow. It also makes LoRA modules easier to store, load, exchange, or disable than full adapted models. At the same time, LoRA is still a training-based personalization method. It requires suitable reference images and training decisions, and it does not by itself solve prompt control, spatial structure, or the risk that the model learns irrelevant details from the reference set (Zhang et al., 2025; Wei et al., 2025).

The contrast between LoRA and IP-Adapter illustrates the practical difference between learned personalization and reference conditioning. In a LoRA workflow, personalization changes or augments the model's parameters for a learned concept, even if the added parameters are small (Hu et al., 2021; Zhang et al., 2025). IP-Adapter and related reference-adapter methods instead condition generation at inference time using image features extracted from one or more reference images (Ye et al., 2023; Wei et al., 2025). Both approaches may help preserve a subject, identity, or style cue, but they occupy different places in the system design. Fine-tuning creates a reusable personalized model component, while reference conditioning provides more situational guidance for a particular generation (Zhang et al., 2025; Cao et al., 2026).

Personalization therefore solves only part of the control problem. It can make a model respond to a particular subject, identity, or style, but the artist may still need other mechanisms to decide where changes happen, how much of an image is preserved, and how the output can be revised after generation (Zhang et al., 2025; Cao et al., 2026; Huang et al., 2025). This is where personalization connects to the broader interface problem discussed in the next section.

1.3 Interface and Interaction Frameworks

The technical capabilities of generative image models only become useful to artists through interfaces that expose those capabilities in workable forms. A generative image workflow may include prompts, reference images, masks, seeds, guidance scales, structural maps, and other conditions (Cao et al., 2026). These inputs become practically useful only when the interface turns them into controls that users can understand, combine, and adjust (Shi et al., 2023a). HCI work on human-generative-AI interaction therefore treats the design space in terms of user purposes, model feedback, user control, engagement level, application domain, and evaluation strategy, rather than only in terms of model capability (Shi et al., 2023a). This framing is especially relevant for image generation because the same model can support very different workflows depending on how its controls are selected and arranged (Shi et al., 2023a; Huang et al., 2025).

The main role of the interface is to translate intent into conditions that the model can use. Text prompting provides broad semantic control over subject matter, style words, attributes, and scene descriptions, but text alone is often insufficient for precise spatial layout, recurring identity, detailed structure, or local preservation (Cao et al., 2026). Controllable generation methods address this limitation by adding visual or structural conditions beyond text (Cao et al., 2026; Zhang et al., 2023). These conditions do not replace prompting. They divide control across modalities: the prompt can describe content and style, spatial controls can specify layout or geometry, and reference images can provide visual, stylistic, or subject-specific cues (Cao et al., 2026; Huang et al., 2025).

Image-editing workflows make this interaction problem more concrete because the user is usually trying to change some parts of an existing image while preserving others. Huang et al. (2025) organize diffusion-based image editing methods by learning strategy, input condition, and editing task, including conditions such as text, masks, reference images, layout, pose, sketches, and segmentation maps. In workflow terms, these methods give users different ways to guide an edit: masks can localize inpainting, an existing image can constrain the new output, reference images can supply visual or stylistic cues, structural inputs can guide pose or layout, and editing instructions can

describe the desired operation (Huang et al., 2025; Cao et al., 2026).

ComfyUI is relevant here because it exposes the conditional pipeline as a configurable node graph rather than hiding it behind a single form. It functions both as an interface and as an inference environment for building generative AI workflows. In image generation, this means that stages such as model selection, adapter modules, encoding, sampling, and decoding can be represented separately (Comfy-Org, 2026). This makes the workflow easier to inspect and reconfigure, since components can be separated, reordered, or replaced through the graph interface rather than only through changes to source code (Comfy-Org, 2026; Abuzuraiq and Pasquier, 2025). This visibility is useful, but it also asks users to translate creative intentions into technical choices about which node, parameter, or conditioning signal to adjust. This reflects a broader HCI finding that many generative-AI interfaces give users direct controls, such as prompts, sliders, or selection tools, while still requiring them to learn how those controls affect the generated output (Shi et al., 2023a). A graph can therefore improve configurability and transparency while still increasing the amount of system logic that users need to understand.

The interface problem is not simply the number of available controls, but the difficulty of knowing which control should change after a particular output. For example, in image-editing workflows, an unwanted result may come from the prompt, mask, reference image, sampling settings, structural condition, or the interaction between them, while the user is also trying to preserve selected parts of the source image (Cao et al., 2026; Huang et al., 2025). Control therefore depends partly on whether the user can connect a possible adjustment with the visual change it is likely to produce (Shi et al., 2023a). Responsiveness affects whether these controls can be used iteratively. Slow generation can make repeated inspection and local refinement harder, while faster sampling, distillation, quantization, and edge-oriented optimization can shorten the iteration loop (Fan et al., 2025; Zeng et al., 2025; Ham and Park, 2026). The next section turns to prompt mediation as a narrower language-level response to this interaction problem, focusing on how LLMs can adapt user prompts before image generation or editing rather than controlling the full workflow (Hei et al., 2024).

1.4 Prompt Mediation for Image Generation

In text-to-image systems, the prompt is often the main interface through which users express what they want the model to generate. Prompt wording is therefore both an interaction problem and a technical one, since different phrasings, keywords, style terms, and levels of detail can change the resulting image (Liu and Chilton, 2022; Xie et al., 2023). The technical reason is that the prompt is encoded into text-conditioning

representations for the image model rather than interpreted as user intention directly. As a result, changes in wording and specificity can affect how linguistic concepts are aligned with visual features during generation (Xie et al., 2023; Cao et al., 2026).

Prompt conventions are also partly model-specific. In many Stable Diffusion-family workflows, prompts are commonly written as compact combinations of subject terms, style words, quality modifiers, and other visual descriptors (Xie et al., 2023). By contrast, official FLUX prompting documentation emphasizes more structured natural-language descriptions that combine subject, action, style, and context (Black Forest Labs, 2026d,e). These model-specific prompting quirks are common enough that several studies address them with systems that take a user's prompt as input and rewrite or expand it before image generation, aiming to improve the generated image while retaining the user's original request (Hao et al., 2023; Hei et al., 2024).

One limitation of specialized prompt-enhancement systems is that they may be tied to a particular model, dataset, or prompting convention (Hao et al., 2023; Hei et al., 2024). Since image models and their preferred prompting styles change quickly, the design response used here treats a general instruction-following language model as a bounded prompt-mediation layer rather than as a fixed prompt optimizer. In this narrower role, the LLM does not need to function as an autonomous creative agent or an open-ended chatbot. It can instead be constrained to reformulate, expand, or vary the artist's own prompt while keeping the user's stated intention as the starting point.

Evidence from ChatGPT-assisted creative tasks should be treated as adjacent rather than direct evidence for image-generation workflows, because these studies mostly examine textual idea generation or writing tasks. Still, they suggest a relevant design concern: LLM assistance can improve the rated quality or creativity of individual outputs, while repeated reliance on similar model suggestions may also make users' ideas more similar at the collective level (Lee and Chung, 2024; Doshi and Hauser, 2024; Meincke et al., 2025). A bounded prompt-mediation design may limit, but cannot remove, the risk that the system normalizes different users' ideas toward similar language. Work on generative AI as a creative aid describes useful roles in ideation, brainstorming, inspiration, and variation, especially when the system provides options for the artist to review, select, and develop rather than replacing the artist's own judgment (Chompunuch and Lubart, 2025; Shi et al., 2023b). In this context, an LLM prompt mediator does not have to only structure an existing prompt. It can also be given a more specific brainstorming task, such as proposing alternative scene directions, style formulations, or prompt variants, while the artist decides which suggestions to use and how to develop them.

Some systems use an LLM to coordinate multiple visual models, plan intermediate structures such as layouts, or iteratively refine prompts and draft images toward a final result (Wu et al., 2023; He et al., 2024; Lian et al., 2023; Yang et al., 2023). Other

work stays closer to prompt mediation by using an LLM to ask clarifying questions and construct text-to-image prompts from the resulting dialogue (Baek et al., 2023). These approaches are useful contrast cases, but they point to different design priorities. Agentic or multi-stage systems often aim to improve prompt following or final-image quality, whereas the narrower role considered here treats the LLM mainly as a text-level mediator whose outputs remain inspectable and easy to revise. This distinction may also matter for local deployment, where controllable diffusion systems involve tradeoffs between quality, control, and efficiency, and additional control mechanisms can increase latency or resource demands (Ham and Park, 2026).

1.5 The Landscape of Local Generative AI

The previous sections described technical elements that can be combined into a creative workflow, including image generation, personalization, editing controls, node-based interfaces, and prompt mediation. Running these elements locally rather than through a proprietary cloud service changes the design problem. The workflow has to fit within available hardware, respond quickly enough for iterative use, expose controls without making the system unmanageable, and keep prompts, reference images, and project material inside the user's own environment.

Local feasibility depends on a stack of optimizations rather than on a single breakthrough. Edge-oriented image generation is usually framed as a tradeoff between quality, efficiency, and controllability: higher fidelity, stronger control, and lower latency are all desirable, but improving one can increase the cost of another (Ham and Park, 2026). In a workflow that combines a base image model with LoRA modules, reference images, and structural controls, each added component can increase memory use, inference time, or computational overhead (Ham and Park, 2026). The practical question is therefore not only whether a model can run locally, but whether the full controlled workflow can run at a speed and quality that still supports creative iteration.

Memory is one of the main constraints. Diffusion models can exceed the VRAM available on ordinary consumer GPUs, especially when run in high precision or combined with additional control and personalization modules. Quantization addresses this by lowering numerical precision, which can reduce model size and support more efficient deployment. The tradeoff is that diffusion-specific architectures and sampling processes can make image quality sensitive to quantization error, so effective methods have to reduce size while limiting degradation (Zeng et al., 2025). Local feasibility therefore depends on more than whether a model fits into VRAM. As an example of this kind of method, Li et al. (2025) report that their SVDQuant approach reduced memory usage for FLUX.1-dev, a 12-billion-parameter model, by about 3.5× compared with the BF16

baseline, with limited reported degradation in visual fidelity.

Latency is the other side of the same problem. Diffusion models usually require repeated denoising or refinement steps, so reducing the number of steps can shorten the feedback loop between prompt, output, and revision (Fan et al., 2025). Distillation methods address this by training a student model to approximate a stronger pretrained model with fewer sampling steps, and sometimes with a more compact architecture. The literature still treats the balance between speed and image quality as a tradeoff rather than a solved problem (Fan et al., 2025). A related deployment distinction appears in recent open-weight releases. The FLUX.2 overview presents FLUX.2 [klein] as an open-weight model family with step-distilled variants for high-volume inference and undistilled base variants for fine-tuning, research, and custom pipelines (Black Forest Labs, 2026c). The official Z-Image repository similarly describes Z-Image-Turbo as a distilled 6B model, while the foundation model is positioned for high-quality generation, diversity, fine-tuning, and downstream development (Z-Image Team, 2025). These sources are release documentation rather than peer-reviewed evaluations, but they show how open-weight image models can be packaged around the distinction between adaptable base models and faster deployed variants.

Local deployment is also relevant for reasons that are not only computational. Cloud services can hide installation, hardware, and maintenance, but they also make the user dependent on external providers for model access, pricing, availability, interface decisions, and data handling. For artists, this can matter when working with unreleased projects, client material, or personal reference images. Work on local AI governance frames locally deployable models as a shift toward user control over model weights, runtime, and data handling, with potential benefits for privacy, autonomy, customizability, offline use, and reduced dependence on centralized providers (Sokhansanj, 2025). At the same time, decentralization should not be treated as an automatic ethical solution. The same shift that gives users more control can also weaken provider-level safeguards and make governance more difficult (Sokhansanj, 2025).

Local workflows also depend on software environments that make the underlying pipeline manageable. As discussed above, ComfyUI is relevant here because its node graph exposes parts of the image-generation pipeline that many simpler interfaces abstract away (Comfy-Org, 2026; Abuzuraiq and Pasquier, 2025). In a local workflow, this makes the pipeline easier to inspect, combine, and revise within the same environment. The cost is that this control also shifts configuration work into the local environment, including node dependencies, parameter settings, memory limits, and interactions between multiple conditioning mechanisms.

A local LLM can support the workflow if its task is kept bounded. Here, that role is prompt mediation: rewriting, expanding, structuring, or varying the user's prompt so that it better fits the image model. Recent open models such as Gemma 4 are relevant

because they are framed as efficient, multimodal, and deployable on local or edge hardware (Google AI Edge Team, 2026). Quantization and GGUF packaging further support this kind of deployment by reducing model size and making local inference more practical. For example, Unsloth's Dynamic 2.0 GGUF documentation describes model-specific quantization schemes intended to reduce size while preserving accuracy, with GGUF files usable in common local inference engines (Unsloth, 2026b). In this bounded role, prompt mediation can remain inside the local workflow rather than sending the artist's prompts, project context, or reference material to an external API.

Local deployment is therefore part of the technical background rather than only a later implementation detail. A private workflow does not remove the tradeoffs described above, but open-weight image models, optimization methods, and bounded local LLM use make a personalized workflow that combines image generation, prompt mediation, and editing plausible enough to evaluate. The concrete hardware, model choices, quantization engine, ComfyUI graph, local LLM backend, and internal testing results are left to the methodology and implementation chapters. The next chapter shifts from this technical possibility to the artist-facing context in which such a workflow would be used: the utility gap in current tools, adoption concerns, questions of authorship and ethics, and the conditions under which a private and controllable system might support creative agency.

Chapter 2

Artists and Generative AI

Generative AI tools are often evaluated by the quality of their outputs, but artists' concerns are broader than whether an image looks convincing. The literature reviewed in this chapter shows that artist-facing systems also raise questions about usefulness in real creative work, control over the process, personalization, trust, privacy, and acceptance.

2.1 Creative Cognition and Mental Models

Artists usually encounter generative AI as an interactive tool within a creative process, not as an abstract model architecture. Its outputs and controls can influence what artists notice, compare, reject, or develop next. The system's usefulness therefore depends not only on image quality, but also on whether artists can form a workable understanding of its behavior and use its controls without interrupting their existing process (Shi et al., 2023a).

One reason generative AI can be attractive in creative work is that it can make possible directions visible very quickly. In artist interviews, text-to-image systems are often described as useful in the early stages of work, where they can provide visual references, explore alternatives, support communication around vague concepts, or help with initial idea development (Ko et al., 2023; Johnston and Thue, 2024). This use fits more naturally with ideation than with finished production. A generated image does not need to be a final artwork; it may be something the artist reacts to, rejects, uses to adjust the direction of the work, or keeps as reference material (Ko et al., 2023; Johnston and Thue, 2024). In that role, the system is less a replacement for creative judgment than a fast source of provisional visual material within an ongoing process.

However, the same prompt-based interaction that makes image generation accessible can also limit ideation. Ko et al. (2023) report that several artists described prompting as a constraint on creativity, especially when the intended visual quality was difficult

to verbalize or when no established vocabulary existed for the concept they wanted to explore. A related problem appears in Rajcic et al. (2024)'s study of artists using personalized text-to-image models. Although the system could be used for ideation, the authors conclude that the general prompt interface is oriented more toward production than toward creative ideation. The limitation is not only that prompts can be difficult to write; it is that the system still needs the artist to provide enough of a direction before it can respond usefully.

Even after a direction has been established, generated images can narrow exploration as well as broaden it. In design research, fixation refers to becoming anchored on a prior example or idea in a way that limits exploration of the wider design space. Wadinambiarachchi et al. (2024) examined this problem in a visual ideation task where participants sketched ideas for a chatbot avatar after seeing an initial example. The study compared participants who received no inspiration support, Google Image Search support, or Midjourney support. Compared with the baseline condition, the Midjourney condition led to sketches that stayed closer to the initial example and to a smaller, less varied, and less original set of ideas. Their qualitative analysis further suggests that fixation can arise both during prompt construction and during reaction to generated images. In the Midjourney condition, participants often built prompts around terms from the design task or the example avatar, and the generated images could then become new anchors for subsequent sketches (Wadinambiarachchi et al., 2024). For tool design, this suggests that ideation support should not only generate polished examples, but should also help users vary their starting terms, encounter less finished or more ambiguous material, and treat outputs as material for further exploration rather than as directions to follow (Wadinambiarachchi et al., 2024).

Mental models are useful here in a practical sense. The term refers to artists' working understanding of what the system is doing, how it responds to input, and what kind of control the user has over it. Zhang and Li (2024) report that novice visual artists held varied and sometimes inaccurate understandings of AI generation and training, ranging from a copy-oriented account of recombining training-image fragments to a more pattern-oriented account of learning visual regularities. In the same study, the more copy-oriented understanding was associated with stronger rejection of AI-generated work as mechanical, derivative, or ethically troubling (Zhang and Li, 2024). This suggests that artists' working theories of the system can shape whether they approach generated images as suspicious copies, useful references, or material for further experimentation.

Used as reference material or as a way to explore alternatives, generative AI can fit into creative work as a source of visual variation (Ko et al., 2023; Johnston and Thue, 2024). The same system can interfere with exploration when difficult visual intentions have to be translated into prompts (Ko et al., 2023; Rajcic et al., 2024). It can also narrow later ideas if generated outputs become too strong as anchors

(Wadinambiarachchi et al., 2024). This tension leads into the rest of the chapter: questions of utility, adoption, authorship, ethics, and privacy all depend partly on whether artists experience the system as supporting their own process or displacing it.

2.2 The Utility Gap in Professional Creative Tools

The previous section framed image generators as useful sources of reference material, variation, and early visual exploration. For professional use, the mismatch is about where support is needed. Popular tools are often built around polished image outputs, whereas many artists are more interested in help around the process: finding references, developing ideas, trying alternatives, and getting feedback (Johnston and Thue, 2024). In this view, the generated image is useful less as an endpoint than as material within an ongoing process. Artist interviews show a similar pattern, with text-to-image models described as useful for reference search, early planning, fast alternatives, and communication with clients or collaborators when concepts were still vague (Ko et al., 2023). This does not mean that finished-image uses are irrelevant. Other work describes cases where AI output becomes a foundation for later refinement, or even a final submitted product, especially in work-oriented contexts (Zhang and Li, 2024). The utility gap is therefore not that image generators produce no useful material, or that final-output use never occurs. It is that their most useful role in many artist-facing accounts is provisional, while popular tools are commonly presented around final-image production (Johnston and Thue, 2024).

A major source of this gap is the prompt interface. Prompt-based systems are accessible because they accept ordinary language, but they also ask artists to translate visual intentions into text. Artists describe this as limiting when the desired concept has no established name, when subtle visual qualities are difficult to verbalize, or when the creative process depends on non-verbal interaction with materials (Ko et al., 2023). Work on personalized text-to-image models makes a related point: a general prompt interface can support ideation, but it remains better aligned with asking for an output than with developing an idea through open-ended visual exploration (Rajcic et al., 2024). This is not only a problem of weak prompts. It also reflects a broader tendency for creative AI systems to remain text-heavy and result-oriented, giving users outputs without enough support for expressing visual intentions, inspecting alternatives, or controlling the process between input and result (Peng, 2024; Shi et al., 2023a).

Professional usefulness also depends on task constraints that are not visible in the image alone. Artists have questioned whether generic text-to-image systems can incorporate requirements such as user flow in interface design, constructability in architecture, or other domain knowledge that exceeds surface appearance (Ko et al.,

2023). These examples show why output quality is an incomplete criterion. A plausible image may still be unusable if it ignores the materials, production constraints, client requirements, or standards of the task. Domain-specific pipelines can address part of this problem by conditioning generation on relevant inputs. A small ComfyUI-based garment pipeline, for example, uses fabric texture inputs and prompts to generate garment images (Ghori et al., 2025). That work is best treated as a proof-of-concept rather than evidence of adoption by designers, but it illustrates the kind of move needed for professional utility. Generation has to be tied to the constraints and materials of a concrete practice.

Personalization does not solve the gap by itself. Fine-tuning on an artist's portfolio can reproduce recognizable fragments such as color, texture, shading, or mark-making, but this does not mean that artists experience the result as an adequate expression of their style (Porquet et al., 2024). In qualitative work with illustrators, style is not a detachable surface that can be transferred onto arbitrary content. It is tied to subject matter, judgment, process, taste, and the artist's way of making decisions (Porquet et al., 2024). A personalized model can therefore be visually faithful in a narrow sense while still failing to support the artist's own practice. The same work makes the stronger argument that style transfer can become a tool of commercial extraction rather than a creativity support tool. That claim is a discussion-level interpretation, but it usefully marks the difference between formal resemblance and practical utility (Porquet et al., 2024).

The studies also suggest ways to bridge the gap. Text-to-image tools can support different levels of variability, domain-specific customization, multimodal input, and prompt assistance, rather than assuming that one text prompt and one output style fit all artistic tasks (Ko et al., 2023). Other artist-facing work points toward reference gathering, ideation, variants, and feedback while preserving the artist's control over important decisions (Johnston and Thue, 2024). Work on expressive interaction adds a related direction: multimodal prompt decomposition, using inputs such as images, colors, and semantic components, can help designers express intentions that are difficult to reduce to text alone (Peng, 2024). These proposals differ in evidence strength, but they converge on the same design implication. Bridging the utility gap requires more than better default images; it requires interfaces that let artists move between rough ideas, references, constraints, generated material, and later revision.

This gap is where the workflow developed here is positioned. Brainstorming, rapid variation, and communicative prototyping are important uses, and in some contexts they may be enough to justify adoption (Ko et al., 2023; Chompunuch and Lubart, 2025). The limitation, for the kind of workflow evaluated in this thesis, is that usefulness depends on more than generating a new image from a prompt. As a design implication, the workflow also needs to support preservation of composition, task constraints, stylistic

decisions, and revision without repeatedly starting over (Ko et al., 2023; Porquet et al., 2024). A workflow that combines image generation with structural control, multimodal guidance, and editing or transformation tools is therefore more relevant here than a prompt-only image generator. Its value is not only in producing images, but in whether generated material can remain editable, situated, and subordinate to the artist's continuing process.

2.3 Adoption Dynamics and Creative Agency

Artists' responses to generative AI are not well described as simple enthusiasm or simple refusal. Artists may recognize practical benefits, including faster iteration, idea development, reference material, feedback, or help with repetitive tasks (Johnston and Thue, 2024; Shi et al., 2023b; Zhang and Li, 2024). At the same time, these benefits sit alongside concerns about ethics, authorship, effort, replacement, and loss of creative agency (Johnston and Thue, 2024; Shi et al., 2023b; Zhang and Li, 2024). Openness to support therefore does not amount to unconditional acceptance of current systems. The important boundary is whether the system extends the artist's process or appears to bypass it (Johnston and Thue, 2024; Zhang and Li, 2024).

Pressure is one source of this conditional acceptance. Shi et al. (2023b) describe this adoption dynamic as an "N-person Prisoner's Dilemma", in which artists feel pushed to use generative AI in order to remain competitive, even when they are uncomfortable with it. In fields that reward rapid iteration, speed and efficiency make the tools difficult to ignore, even when concerns about copyright, user experience, and market effects remain unresolved (Shi et al., 2023b). Qualitative work with novice visual artists shows a related developmental pattern, where initially negative views can become more ambivalent after artists encounter the speed and convenience of AI tools, without resolving concerns about originality, effort, or quality (Zhang and Li, 2024). Adoption can therefore be pragmatic rather than enthusiastic.

Creative agency is a central condition that separates acceptable assistance from threatening automation. Artists are more willing to use systems that leave creative judgment intact and more resistant to systems that appear to replace ideation, decision-making, or expressive labor (Johnston and Thue, 2024; Zhang and Li, 2024). This helps explain why support for references, variants, critique, unwanted tasks, or later refinement is often easier to accept than tools framed around finished image production (Johnston and Thue, 2024; Zhang and Li, 2024). The issue is not just interface control. It is whether the artist can still recognize the work as emerging from their own intentions, taste, and effort (Johnston and Thue, 2024; Zhang and Li, 2024).

This concern with agency is also tied to how artistic labor is valued. Artists do

not judge AI-assisted practice only by output quality or speed. They also consider which parts of the process the system removes, and whether the remaining human contribution is enough for the work to feel legitimate (Johnston and Thue, 2024; Zhang and Li, 2024). One source of rejection is the feeling that AI-assisted creation can be "too easy", making the result feel detached from the artist's own investment in the work (Zhang and Li, 2024). At the same time, artists show interest in support for unwanted, repetitive, or physically difficult tasks when this does not take over the decisions through which the work becomes theirs (Johnston and Thue, 2024; Zhang and Li, 2024). Conceptual discussions connect these concerns to authenticity, the value of human-made art, and the place of human intention in artistic practice (Jiang et al., 2023; Garcia, 2025). Those broader claims are interpretive rather than direct empirical findings, but they fit the interview evidence in which labor, authorship, and agency are difficult to separate (Johnston and Thue, 2024; Zhang and Li, 2024).

Attitudes also vary across artists, domains, and levels of understanding. In Johnston and Thue (2024), a notable minority of artists remained unsure about AI assistance, which suggests that some positions are still unsettled rather than firmly polarized. Shi et al. (2023b) further show that the same technology may be interpreted as a tool, collaborator, or competitor depending on the kind of work at stake. Mental models can also shape artists' responses: incomplete or inaccurate understandings may intensify discomfort, while more differentiated understandings can support more nuanced judgments (Zhang and Li, 2024). Adoption is therefore conditional on what artists are trying to do, what risks they associate with the tool, and how much control they feel they retain.

Usefulness alone is therefore not enough. Generative AI may be more acceptable when it supports ideation, exploration, feedback, or unwanted and repetitive work, but resisted when it compresses the creative process into rapid output production (Johnston and Thue, 2024; Shi et al., 2023b; Zhang and Li, 2024). It may also be resisted when practical benefits are tied to unresolved concerns about consent, authorship, and the use of artists' work in training data (Johnston and Thue, 2024; Shi et al., 2023b; Zhang and Li, 2024). This makes adoption dynamics and creative agency difficult to separate from the ownership and ethics issues that follow, including contribution, credit, consent, and control over artistic material.

2.4 Ownership, Authorship, and Ethics

Ownership and authorship are difficult to assign in generative AI contexts because an image can reflect several kinds of contribution at once. This becomes especially visible when a system produces images in a recognizable artist's style. In that situation,

the output may depend on prior artwork, model training, prompting, choosing among generated outputs, and sometimes further editing. In the survey reported by Lovato et al. (2024), similar proportions of respondents agreed that ownership should belong to the person using the model or to the artist whose style was represented, while fewer supported ownership by the model creators. The result does more than show disagreement about legal detail. It points to uncertainty about where authorship should be located when style, training data, and user action are entangled (Lovato et al., 2024; Garcia, 2025).

Survey evidence suggests strong support for detailed disclosure of training data (Lovato et al., 2024). On compensation, about half of respondents selected options where they did not need personal profit but cared who profited from their work; the most common single response opposed for-profit companies profiting from their art (Lovato et al., 2024). Where training datasets or searchable subsets have been available, artists have reported finding their work included despite not having consented or been credited; for closed systems, the lack of dataset disclosure makes such inclusion harder to verify (Jiang et al., 2023). Style-mimicry tools can also attach an artist's recognizable style to outputs they did not approve (Jiang et al., 2023). This creates risks of reputational damage, lost credit, economic harm, and reluctance to share work publicly (Jiang et al., 2023; Garcia, 2025). In this sense, the ethical concern is not just copying, but the treatment of artistic labor as an extractable input for systems that artists may not control (Jiang et al., 2023; Garcia, 2025).

When AI is described as personalized to the user's own prior artwork, observers assign more credit to the human user than when the same images are described as produced with standard AI (Khan et al., 2025). However, the same increase in perceived human contribution did not carry over to aesthetic appreciation or to willingness to describe the images as "true art" (Khan et al., 2025). Related experimental work also suggests that cues of artist involvement matter, since training an AI system on curated personal images and explaining the artist's intended vision increased perceived authenticity in AI-assisted artworks (Messer, 2024). Both findings concern lay judgments rather than artists' own experience of using personalized systems. Together, they suggest that artist-specific cues can make human involvement more legible, while still leaving unresolved broader judgments about originality, authenticity, and artistic value (Khan et al., 2025; Messer, 2024; Garcia, 2025).

These issues affect artists in practical and emotional ways, not only in legal terms. In a small qualitative study, Kambur and Dolunay (2024) report that weak copyright protection and uncertainty around AI-generated works were described as harmful to artists' emotional states, with possible effects on motivation and willingness to create. The evidence is limited, but it aligns with the broader concern that artists lose recognition and control when their work or style can be reused without consent or clear

authorship (Jiang et al., 2023; Garcia, 2025).

Ownership and ethics affect whether artists are likely to experience generative systems, especially personalized ones, as supportive or extractive. Rights assignment is part of the issue, but the practical concern is broader: whether the system respects consent, keeps human contribution visible, and avoids turning an artist's recognizable style into a detached resource for others to use (Lovato et al., 2024; Khan et al., 2025; Jiang et al., 2023; Garcia, 2025). A private and artist-controlled workflow can respond to some of these concerns by keeping training material, prompts, and outputs closer to the artist, but it does not by itself settle questions of consent or authorship.

2.5 Privacy and Decentralized Infrastructure

In artistic domains, privacy is tied to more than the confidentiality of prompts or uploaded files. It also concerns unauthorized style mimicry, dataset inclusion without consent, and loss of control over how creative work is reused (Lovato et al., 2024; Jiang et al., 2023). Privacy therefore overlaps with ownership and authorship when artists' work enters generative systems and can circulate beyond their control (Jiang et al., 2023; Shan et al., 2023).

Concerns about style mimicry and dataset reuse have also prompted technical defenses. Glaze, for example, is designed to make publicly shared artworks harder to use for later style mimicry by applying small perturbations before they are posted online (Shan et al., 2023). In the reported evaluations, the method frequently reduced successful mimicry attempts while remaining visually tolerable for many artist participants (Shan et al., 2023). For detecting reuse, DIAGNOSIS coats protected images with subtle signals that can later be detected in models trained or fine-tuned on that data, making unauthorized use easier to identify after the fact (Wang et al., 2023). Together, these papers show that style mimicry and dataset reuse are being treated as technical problems as well as legal or policy problems (Shan et al., 2023; Wang et al., 2023).

The same issue appears before work is public, when artists are still using prompts, references, training material, and drafts inside a workflow. Running models on user-controlled hardware can reduce dependence on cloud providers and keep more of that process under the user's control (Kabir et al., 2024; Sokhansanj, 2025). This is relevant to artists in a practical sense: prompts, reference images, training material, and unfinished project files may be commercially sensitive, personally meaningful, or tied to client work. A local workflow cannot prevent unauthorized reuse of work that is already public, but it can reduce the need to send new material through external services during ideation, personalization, and editing.

Local deployment is also technically plausible in a way that would be difficult

if high-quality image generation required large centralized infrastructure. Consumer-hardware image generation has been demonstrated as an implementation direction, while the optimization literature describes several ways to reduce inference cost, including distillation, quantization, and edge-oriented model design (Kabir et al., 2024; Fan et al., 2025; Zeng et al., 2025; Ham and Park, 2026). This does not mean that local workflows are easy for artists to install, maintain, or integrate into professional practice. It means that local image generation is feasible enough to evaluate as a design option rather than only discuss as a privacy ideal.

Privacy is only one part of the control argument. Local deployment by itself does not guarantee transparency, and a cloud-based system could still expose some controls. The stronger design possibility comes from combining local execution with an inspectable workflow. In ComfyUI-based work on explainability for artistic practice, exposing and manipulating model components is framed as a way to support artistic agency and tacit understanding (Abuzuraiq and Pasquier, 2025). Broader HCI work similarly identifies hidden intermediate model processes as a limitation in many generative-AI systems (Shi et al., 2023a). This matters because control over a generative system is not only a matter of picking the best final output. In design research, some creative professionals wanted low-effort steering, while others wanted more precise input control or less automatic interpretation by the system (Peng, 2024).

What appears less developed, in the literature reviewed here, is direct artist-facing evaluation of local or private generative workflows. Existing work gives strong reasons why artists may care about control, consent, reuse, authorship, and provider dependence (Lovato et al., 2024; Jiang et al., 2023; Johnston and Thue, 2024; Shi et al., 2023b). Technical and governance work suggests that local deployment can shift some control over data handling, model access, and runtime environment back toward the user (Kabir et al., 2024; Sokhansanj, 2025). There is less evidence on whether artists experience that shift as useful in practice, what burdens it introduces, and how privacy interacts with creative utility, personalization, and control during actual use.

A local, inspectable workflow is therefore worth evaluating because it addresses several artist-facing concerns at once. It can reduce exposure of unpublished material, make personalization less dependent on external providers, and turn model choice, adaptation, conditioning or sampling into explicit creative decisions rather than hidden service decisions. In that sense, privacy, personalization, transparency, and creative control become part of the same design problem (Kabir et al., 2024; Sokhansanj, 2025; Abuzuraiq and Pasquier, 2025; Peng, 2024).

Chapter 3

Methodology

This chapter describes how the empirical material used in the thesis was produced and analyzed. It covers the research design, evaluation framework, participant recruitment, procedure, data collection, analysis strategy, and methodological limitations. Detailed technical implementation of the workflow versions is left to Chapters 4 and 6, while evaluation-specific findings are presented in Chapters 5 and 7.

The empirical work is treated as exploratory qualitative evaluation rather than as controlled experimentation. Its aim is to develop a transparent account of how the workflow was experienced in use and how those observations informed design decisions.

3.1 Research Design and Evaluation Framework

The empirical part of the thesis was designed as a small exploratory evaluation of a working design artifact: a private, stylistically personalized generative AI workflow for artists. This design was chosen because the thesis concerns situated use rather than controlled performance measurement. The evaluation therefore asks how artists experienced the workflow, what kinds of creative support it offered, where it broke down, and how its private and personalized character affected trust, control, and acceptance.

With a small number of participants and a novel workflow, the purpose was to produce a detailed account of interaction with the artifact and to identify implications for design. The overall thesis has an iterative design logic: an initial workflow was built, evaluated with artists, analyzed for strengths, limitations, and design implications, and then used to motivate a redesigned workflow. The redesigned workflow was then evaluated in a second round of supported sessions. The two rounds are treated as linked exploratory evaluations rather than as controlled comparisons.

The evaluation framework was organized around four objectives derived from the aim of the workflow. These objectives served two roles. First, they guided the design of the study protocol, questionnaire, and interview. Second, they provided the initial

deductive frame for later analysis. The objectives are summarized in Table 3.1.

Objective	**Guiding question**	**Main data sources**	**Example indicators**
Usability	Is the workflow easy to learn and operate in the study setting?	Screen recording, transcript, debrief interview, questionnaire	Expressions of confusion or ease, requests for help, errors, comments on complexity
Creative utility	Do artists find the workflow creatively useful?	Screen recording, participant comments, generated images, debrief interview	Expressions of inspiration, use of generated images as references, comparisons with existing creative methods
Personalization and stylistic fidelity	Do participants recognize their own visual style or specific stylistic regularities in the outputs?	Generated images, transcript, debrief interview, questionnaire	Self-recognition comments, captured or missed stylistic features, ratings of style match
Control, trust, privacy, and acceptance	Do artists feel comfortable, in control, and confident about how their material is handled?	Debrief interview, questionnaire, transcript	Statements about ownership, data security, local processing, trust or distrust, willingness to use the workflow

Table 3.1: Evaluation framework used to guide data collection and analysis.

The analysis was framework-analysis-informed. Framework analysis is useful for applied qualitative work because it supports systematic organization of data while still allowing comparison across cases and themes (Gale et al., 2013). In this thesis, the four evaluation objectives provided the initial deductive structure. At the same time, the analysis remained open to more specific themes that were not fully anticipated in the protocol but were relevant to the participants' interaction with the workflow. This hybrid deductive-inductive approach was used to keep the analysis aligned with the evaluation objectives without forcing all observations into predefined categories.

The study also included a post-session questionnaire. Its usability section was based on the System Usability Scale, a short standardized scale intended for global assessment of perceived usability (Brooke, 1996). Additional questionnaire items were written for the project-specific objectives of creative utility, personalization, control, privacy,

and trust. These questionnaire responses are used descriptively to contextualize the qualitative findings, not as statistical evidence.

3.2 Participants and Recruitment

For the first evaluation, participants were selected purposively according to the needs of the study and recruited through convenience-based access to the researcher's personal and academic networks. The purpose was not to obtain a representative sample of artists, but to recruit participants whose creative practice made the workflow possible and meaningful to evaluate.

The inclusion criteria were both practical and methodological. Participants needed at least one year of sustained or professional creative experience, a portfolio containing at least 15 works suitable for participant-specific personalization, and willingness to share images for training a custom LoRA model before the session. The portfolio requirement was necessary because the workflow depended on stylistic personalization; without enough participant-specific material, the evaluated artifact could not be instantiated in the intended form.

Two participants were recruited for this first evaluation. Participant 1 (P1) was an illustrator and digital artist recruited through the researcher's personal network. Participant 2 (P2) was a 3D mapping, animation, and VJ-style practitioner recruited through the researcher's academic network. They had more prior experience with generative visual AI tools in creative work.

The redesigned workflow evaluation included P1 again and one additional participant, Participant 3 (P3). P1's return makes their second session a repeat participant case, while P3 provides an additional perspective on the redesigned artifact. P3 was a digital artist and student with experience in node-based creative tools and image-based digital work, including tools such as TouchDesigner and Blender. They also had some prior experience with local language-model systems. Their prior use of generative image tools was limited and only involved cloud-based services.

All participants gave informed consent for academic use of the session recordings and for the inclusion of generated or edited images and portfolio images provided for LoRA training in the final thesis.

Across both evaluation rounds, P1 was known personally to the researcher and had passive familiarity with the project and workflow development. This is treated as a reflexive limitation later in the chapter.

3.3 Procedure and Data Collection

The first evaluation concerned the first version of the local generative image workflow built in ComfyUI. For each participant, the workflow was personalized before the session by training a custom LoRA model from images provided from their portfolio. After training, the model was inspected through test generations to check whether recognizable stylistic patterns from the participant's material appeared in the outputs. The full technical design of the original workflow, including the model stack and personalization pipeline, is described in Chapter 4.

Each session in the first evaluation followed the same broad structure, although the timing and emphasis were adjusted to the participant's pace and interests. At the beginning of the session, the researcher introduced the purpose of the evaluation, confirmed consent, and started the screen and audio recording. The participant was then given access to the workflow through the ComfyUI interface. The researcher explained the main parts of the interface. After this walkthrough, the participant had a short period of free exploration to become familiar with the workflow.

The main part of the session was a flexible task-based exploration. Participants could choose between several predefined creative brief topics, work on an existing personal or professional project, or use the workflow for open-ended ideation. The prepared task options are reproduced in Appendix 8.6. Although predefined topics were available, they functioned mainly as fallback prompts. In practice, all sessions developed into open-ended ideation or participant-supplied creative problems rather than completion of a standardized task. Once a task direction had been chosen, participants were encouraged to use the workflow freely, react to outputs, change prompts or parameters, and follow emerging interests. The planned protocol therefore provided a structure, but the researcher adjusted the pace when participants needed more explanation or moved more quickly than expected.

The researcher remained nearby throughout the session. Their role was to introduce the workflow, answer participant questions when asked, provide technical support if needed, and monitor the system. Follow-up questions were asked occasionally when participants reacted to an output or described uncertainty. The sessions should therefore be interpreted as researcher-supported evaluations rather than as unattended independent use.

After the exploratory part of the session, participants took part in a semi-structured debrief interview and completed the post-session questionnaire described in the previous section. All sessions were conducted in English except P2's session, which was conducted primarily in Slovak because that was more comfortable for the participant.

The material from the first evaluation consisted of screen and audio recordings of the full sessions, automatically generated transcripts, all images generated during the

sessions, and post-session questionnaire responses. P1 generated 119 images and P2 generated 150 images during their sessions. Together, these materials formed the basis for the qualitative analysis described in the following section.

The redesigned workflow evaluation followed the same general format as the first evaluation sessions. The main change was the artifact: participants used the redesigned split generation-and-editing workflow described in Chapter 6. As in the first evaluation, the researcher remained available for explanation and technical support, so the redesigned workflow was evaluated as a case of use with researcher support rather than as independent deployment. The material from the redesigned workflow evaluation consisted of screen and audio recordings, transcripts, generated and edited images, debrief material, and post-session questionnaire responses. In the redesigned workflow sessions, P1 generated 82 images and P3 generated 96 images.

3.4 Data Analysis

The analysis developed in two stages. The first stage took place shortly after data collection and was based on the researcher's direct observation during the sessions, review of generated images and questionnaire responses, and selective checking of transcripts and screen recordings around salient moments of use. This stage identified patterns for each participant, important interaction episodes, and preliminary design implications. It was not an exhaustive line-by-line coding of the full session material, but it was an analytic pass in which observed moments were interpreted in relation to the evaluation objectives.

The second stage made the interview analysis more explicit and traceable through a matrix-based organization of the debrief interviews. A coding unit was treated as a meaningful response segment rather than a fixed sentence or turn of talk. Each segment was assigned to one or more of the four evaluation objectives: usability and learnability, creative utility, personalization and stylistic fidelity, and control, trust, privacy, and acceptance. More specific descriptive themes and short analytic memos were recorded at the same time where they helped characterize the participant's response more precisely.

The matrix included separate entries for participant, transcript location, excerpt or paraphrase, evaluation objective, descriptive theme, and interpretive note. This structure was used to prepare within-case summaries for each participant before comparing patterns and contrasts across the relevant cases.

The screen recordings and generated images were not coded in the interview matrix. They were used as contextual evidence for interpreting participant comments and selected workflow episodes, especially when a participant referred to a specific output, difficulty, or moment of use. Generated images are not treated as independent outcome measures,

and questionnaire responses are used descriptively to support and contextualize the qualitative findings.

AI tools were used as a supplementary aid in organizing selected transcript and session material into a working overview of possible themes and relevant moments. This AI-assisted overview is treated as an analytic aid and audit-trail document, not as the primary source of findings. Claims made in the thesis are based on manual interpretation of the debrief interviews, selected session episodes, questionnaire responses, and generated images. The researcher remained responsible for interpreting the material and deciding which evidence to include.

All sessions were analyzed first as individual participant cases using the same evaluation objectives. The first evaluation was then synthesized across P1 and P2. The redesigned workflow evaluation uses the same general logic for P1 and P3. Because P1 participated in both rounds, findings from P1's first session were also used as comparative context when interpreting their redesigned workflow session.

3.5 Reflexivity and Methodological Limitations

Several features of the study shape how the findings should be interpreted. The researcher designed and configured the workflow, introduced it to the participants, recruited the participants, conducted the interviews, and analyzed the material. This position gave the researcher detailed knowledge of the artifact and made it possible to connect observed interaction problems to later redesign decisions. At the same time, it means that the evaluation was not independent of the design process. Participant questions, researcher explanations, and technical support were part of the study setting and need to be considered when interpreting the findings.

The findings therefore describe use in a researcher-supported exploratory setting. Participants were not asked to install, train, troubleshoot, or maintain the workflow independently. The study can speak to how artists interacted with the workflow when it was available in this setting, but it should not be read as evidence that the same workflow could be adopted without support. This distinction is especially important because local personalized AI workflows involve technical work beyond ordinary use, including model training, hardware constraints, dependency management, and parameter tuning.

The researcher's personal familiarity with P1 also shaped the origin of the project. Early informal discussions around P1's creative practice and concerns about generative AI helped motivate the focus on personalized, artist-centered tools. The research questions, evaluation objectives, and initial workflow design were then developed through literature review, technical exploration, and implementation work, while the later redesign was also informed by the first evaluation. This background is relevant

reflexively because it was one source of motivation and may have shaped which design problems the researcher noticed and prioritized.

The recruitment process also shaped the data. The sample was small, purposive, and based on access to artists whose portfolios and practices made the workflow possible to evaluate. This allowed the evaluation to focus on information-rich cases, but it limits the extent to which the findings can be generalized beyond these participants and sessions. For P1 specifically, across both evaluation rounds they were known personally to the researcher and had passive familiarity with the project and workflow development, including some familiarity with the redesigned workflow. Their second session was also shaped by prior exposure to the first artifact and the earlier evaluation setting. These factors may have affected comfort, trust, willingness to criticize the workflow, or ease of interaction with the researcher.

Across the evaluations, sessions differed in participant trajectory, prior exposure to the workflow, and, in one case, language. All sessions were conducted in English except P2's first-evaluation session, which was conducted primarily in Slovak because that was more comfortable for the participant. Quotations from English-language sessions are reproduced from the original English transcript after checking against the recordings. Quotations from P2's Slovak-language session are translated into English by the researcher. Automatic transcripts were treated as working materials rather than final records, particularly in moments where speaker attribution or wording was uncertain. Quotations included in the results were checked against the original recordings.

The questionnaire data were also interpreted cautiously. The usability items based on the System Usability Scale provide a useful descriptive reference point, and the additional items help summarize participant responses to creative utility, personalization, control, privacy, and trust. With this small sample, these results are used only to contextualize the qualitative interpretation.

Chapter 4

Original Workflow: Design and Implementation

4.1 Design Goals and Workflow Overview

The first workflow was designed as a private, stylistically personalized image-generation environment for artists. Its intended role was to produce visual material that participants could inspect, reinterpret, and use as reference within their own creative process, rather than to automate the production of finished artworks. This framing followed from the research aim: the workflow was meant to test whether local deployment and participant-specific personalization could make generative image output more useful, trustworthy, and personally relevant for artists.

The design therefore combined four requirements. First, the workflow had to keep the participant's portfolio images, prompts, trained model files, and generated outputs inside a controlled local environment rather than sending them to a public cloud service. Second, it had to support stylistic personalization, so that the generated images could reflect visual regularities from the participant's own work. Third, it had to remain flexible enough for exploratory use, where the participant could change prompts, compare outputs, and use image references while following their own interests during the session. Fourth, it had to be usable in a supported study setting by artists who were not expected to have specialist knowledge of diffusion models or local AI infrastructure.

The resulting artifact combined a local ComfyUI workflow, a high-quality open-weight image model, participant-specific LoRA modules, and optional image-conditioning paths. These components were assembled as a single workflow that could be prepared before a session and then used with a participant through a browser interface. The detailed implementation of the deployment setup, generation stack, personalization pipeline, and interface is described in the following sections.

The initial workflow deliberately kept the interaction scope narrower than the full

Component	Implementation in the original workflow
Deployment setup	Dedicated local workstation for inference and model files; researcher-provided laptop used as the participant-facing browser client.
Hardware	NVIDIA RTX 4080 Super GPU on the workstation.
Interface and inference environment	ComfyUI node graph.
Base image model	FLUX.1 [dev].
Inference optimization	Nunchaku/SVDQuant low-bit inference to reduce memory use and improve responsiveness.
Personalization method	Participant-specific LoRA module trained before the session and loaded into the workflow during use.
Training material	Portfolio subset of 15–50 participant-provided images, depending on available and suitable material.
Captioning and trigger words	Local vision-language-model captioning with a shared uncommon trigger string and participant-specific style keyword.
Generation paths	LoRA-only path, LoRA with FLUX Redux, and LoRA with IP-Adapter, each producing separate outputs for comparison.

Table 4.1: Summary of the main implementation choices in the original workflow.

set of available diffusion controls. This was partly a design decision and partly a practical constraint. The first artifact prioritized personalized reference generation and stylistic exploration, because these uses matched the initial evaluation focus on privacy, personalization, usability, and creative utility. Adding dedicated inpainting, image editing, or stronger structural guidance would also have made the workflow more technically demanding and less likely to remain quick and interactive in the local setup. For this reason, the first version exposed image-conditioning paths, but did not attempt to function as a full editing or transformation interface.

4.2 Private Deployment Setup

The original workflow ran on a dedicated workstation with an RTX 4080 Super GPU, while participants interacted with it from a researcher-provided laptop through a browser-based ComfyUI interface. This arrangement was mainly practical: the workstation

provided the GPU resources needed for local inference, while the laptop provided a portable participant-facing display. The participant-specific LoRA models were trained before the sessions and then loaded into the workflow during use. During the sessions, the laptop functioned only as the client interface; the participant data, model files, and generated outputs remained on the workstation.

This arrangement supported the privacy and control aims of the study in a practical way. Participant portfolio images did not need to be uploaded to a commercial image-generation service, and the trained LoRA files and generated images remained in a researcher-controlled workspace. This allowed the participant-facing interaction to remain browser-based, even though the underlying system still required technical preparation and monitoring by the researcher.

The setup should therefore be understood as a private deployment supported by the researcher rather than as an independently installed consumer application. It tested whether artists could meaningfully interact with a local personalized workflow when it was made available to them, not whether they could install, configure, and maintain the full infrastructure themselves.

4.3 Image Generation Stack

The image-generation stack was organized around ComfyUI, which provided both the local inference environment and the visual node graph used to assemble the workflow (Comfy-Org, 2026). This choice made it possible to combine a base image model, participant-specific LoRA files, image-conditioning modules, and sampling controls in one inspectable workflow. For the purposes of the first evaluation, ComfyUI was useful because the system could be adjusted and monitored by the researcher while still presenting the participant with a browser-based interface.

FLUX.1 [dev] was used as the base image model (Black Forest Labs, 2024). It was selected as a high-quality open-weight model that could be run in the local workstation setup and combined with the personalization and conditioning components needed for the experiment. In the workflow, the participant's prompt was encoded by the model pipeline and used together with the active LoRA and optional image-conditioning inputs to generate image candidates. The system was therefore not only a single text-to-image form, but a configurable pipeline in which text, learned style, and image references could all contribute to the output.

Because FLUX.1 [dev] is a large model, the workflow used Nunchaku to run a low-bit version of the model more efficiently on the available hardware. Nunchaku is the inference engine released with SVDQuant, which was designed to support 4-bit quantized diffusion-model inference while reducing memory use and limiting quality

loss (Li et al., 2025). In the original workflow, the practical role of this optimization was to make the local setup responsive enough for interactive exploration during the evaluation sessions.

The main generation loop followed the structure common to diffusion-based image workflows. The participant supplied or revised a prompt, selected the relevant personalization and image-conditioning controls, and queued a generation. The workflow then produced image outputs, which were shown in the ComfyUI interface and saved on the workstation. Seeds and other sampling parameters could be changed between runs, allowing variations to be explored without changing the whole workflow. This made the stack suitable for reference generation and stylistic exploration, where outputs could be inspected, compared, and used as material for further prompting or manual creative work.

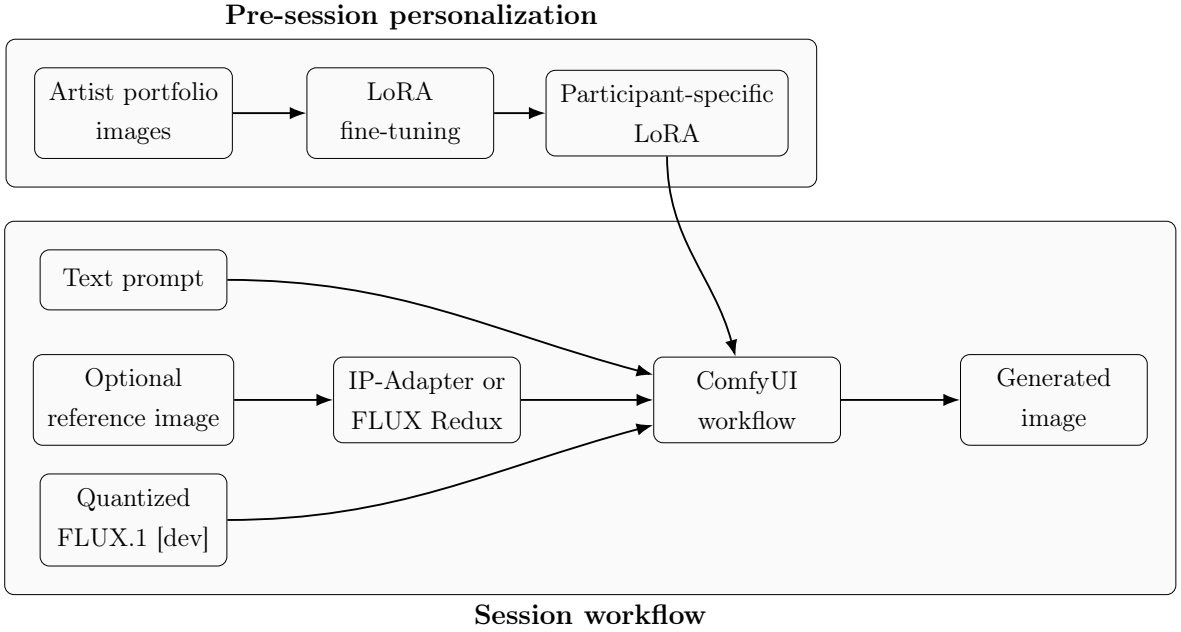


Figure 4.1: Simplified structure of the workflow.

4.4 Stylistic Personalization Pipeline

Stylistic personalization in the original workflow was implemented mainly through participant-specific LoRA modules. LoRA was suitable for this artifact because it allowed the base model to be adapted through a comparatively small set of learned weights rather than through a separate full model checkpoint (Hu et al., 2021). In practical terms, this meant that each participant's style could be represented as a separate file that was loaded into the workflow during their session.

The personalization process took place before the evaluation sessions. Each participant provided a portfolio subset containing between 15 and 50 images, depending on

the available material and suitability for training. These images were prepared as a participant-specific training set. The preparation included captioning the images with a local vision-language-model workflow and assigning activation words that could later be used in prompts to refer to the trained style. In this implementation, each caption used the same uncommon trigger string, `ohxw`, paired with a short style keyword: `lineart` for one participant and `pattern` for the other. The rest of the caption described the visible content of the image without adding general style words. After automatic captioning, the captions were briefly reviewed to correct obvious errors. This was a practical training decision rather than a separately validated methodological contribution. The intention was to give the LoRA a consistent phrase for the participant-specific style while keeping the remaining caption text focused on image content.

The LoRA modules were trained with Kohya-SS sd-scripts, which provide training scripts for LoRA and related diffusion-model adaptation workflows (Kohya-SS, 2026b). During training, checkpoints were saved at regular intervals so that the model could be inspected and the risk of overfitting could be managed. The aim was not to produce a general-purpose model of the artist, but to create a reusable style-conditioning module that made the outputs visibly closer to the participant's own work while still allowing new prompts and image references to guide the generated content.

4.5 Interface and Interaction Design

The participant-facing interface was the ComfyUI graph itself, prepared in advance so that the main controls needed for the session were visible in one workspace. The graph was organized into three main generation paths: a LoRA-only path, a FLUX Redux path, and an IP-Adapter path. Each path produced its own image output, allowing the participant to compare how the same prompt behaved under different forms of image conditioning. All three paths could use the participant-specific LoRA module, with an adjustable LoRA strength. The LoRA-only path emphasized this style module by itself. The Redux path added variation and restyling from an uploaded reference image within the FLUX ecosystem, while the IP-Adapter path added a separate reference-image adapter for image-prompt conditioning.

Several controls were placed near the top of the graph to make the basic generation loop easier to manage. The text-prompt field was the main input for describing the desired image. A seed control allowed the generation either to produce a new variation or to keep the same random seed while other settings were changed. Short note panels explained the prompt, seed, generation toggles, and a known out-of-memory recovery step. These notes were included because the workflow exposed technical controls that would not be self-explanatory to a participant encountering ComfyUI for the first time.

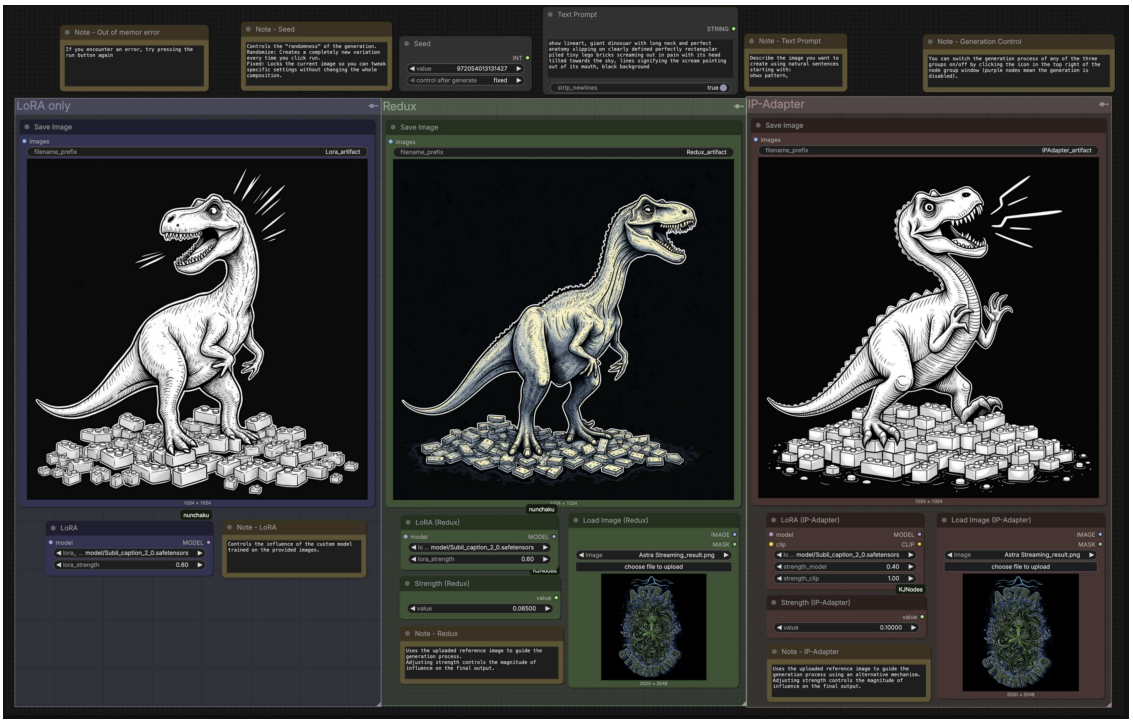


Figure 4.2: ComfyUI interface for the original workflow.

The three output areas were color-coded and spatially separated. This made the workflow legible as a comparison between different generation modes rather than as one large undifferentiated node graph. In the LoRA-only group, the participant could inspect the output produced primarily from the text prompt and the participant-specific LoRA, with a LoRA strength control setting how strongly the trained style module influenced the result. In the Redux and IP-Adapter groups, image-upload nodes and strength controls made it possible to introduce a reference image and adjust how strongly it influenced the output. Each generation path could also be turned on or off with a single toggle, so the participant did not have to run all methods at once. This made it possible to generate a faster baseline output through one path and then activate additional conditioning only when needed. These controls made the workflow more flexible than a prompt-only generator, but they still acted as generation-conditioning controls rather than as direct editing tools. The generated images were shown directly inside the interface and saved with separate filename prefixes, which made it possible to distinguish outputs from the different paths during later review.

4.6 Implementation Constraints and Known Limitations

The main limitations of the original workflow followed from its scope as a local, personalized ComfyUI artifact. The system had to run a large image model on the available

workstation, combine several conditioning mechanisms in one graph, and remain usable enough for a supported evaluation session. These requirements made the workflow feasible to test, but they also created three important boundaries: GPU memory and stability, dependence on researcher setup, and a deliberately narrow interaction scope.

The main practical constraint was the available GPU memory. The RTX 4080 Super made the workflow feasible, especially with Nunchaku optimization, but the graph still combined a large base model with multiple conditioning mechanisms. During technical testing, the first activation of the IP-Adapter path could trigger an out-of-memory error. This behavior was known before the evaluation sessions, and a short note was placed in the interface instructing the user to re-queue the prompt if it occurred. In the sessions, re-queueing resolved the issue and the workflow continued to function, but the episode illustrates that local modular workflows can expose resource-management problems that are often less directly visible in hosted services.

A second constraint concerned maintenance and setup. Participants interacted with the workflow after the environment, model files, custom LoRA modules, and ComfyUI graph had already been prepared. They were not asked to install ComfyUI, manage dependencies, configure GPU inference, or train LoRA modules themselves. This was consistent with the exploratory aim of the study, but it limits what the evaluation can show. The study evaluates supported use of a private personalized workflow, not independent adoption of the full technical stack.

Finally, the artifact was scoped around personalized generation and image-conditioned reference creation. It included image inputs through IP-Adapter and FLUX Redux, but these inputs acted as conditioning signals within generation rather than as a full image-manipulation interface. The workflow did not include dedicated image-editing tools or other specialized image-input processing paths.

Chapter 5

Original Experiment: Evaluation

This chapter presents the results of the first evaluation of the original workflow. The findings are organized by participant before turning to a cross-case synthesis and the lessons that motivated the redesign. As described in Chapter 3, the results are based on researcher-supported exploratory sessions, debrief interviews, questionnaire responses, generated images, and screen recordings. Questionnaire scores are used descriptively, while the main analysis is based on the qualitative session material.

5.1 Results: Participant 1 — The Illustrator

P1 used the workflow in relation to a personal illustration project. In the session, the workflow was used mainly for reference generation, compositional exploration, and possible help during moments of ideation difficulty. P1's questionnaire responses were positive: their System Usability Scale score was 82.5, and their project-specific ratings were high for creative utility, personalization, and control and trust. These scores are used here as descriptive context rather than as standalone evidence.

Metric	**Score**	**Descriptive interpretation**
Usability (SUS)	82.5 / 100	High perceived usability, with prompting and control predictability still described as difficult.
Creative utility	4.4 / 5.0	High perceived usefulness, especially for reference generation and sketch-phase support.
Personalization	5.0 / 5.0	Very strong perceived style match.
Control and trust	5.0 / 5.0	Strong trust in the private setup and positive ownership/control ratings, though ownership was discussed as philosophically uncertain.

Table 5.1: Descriptive questionnaire summary for Participant 1.

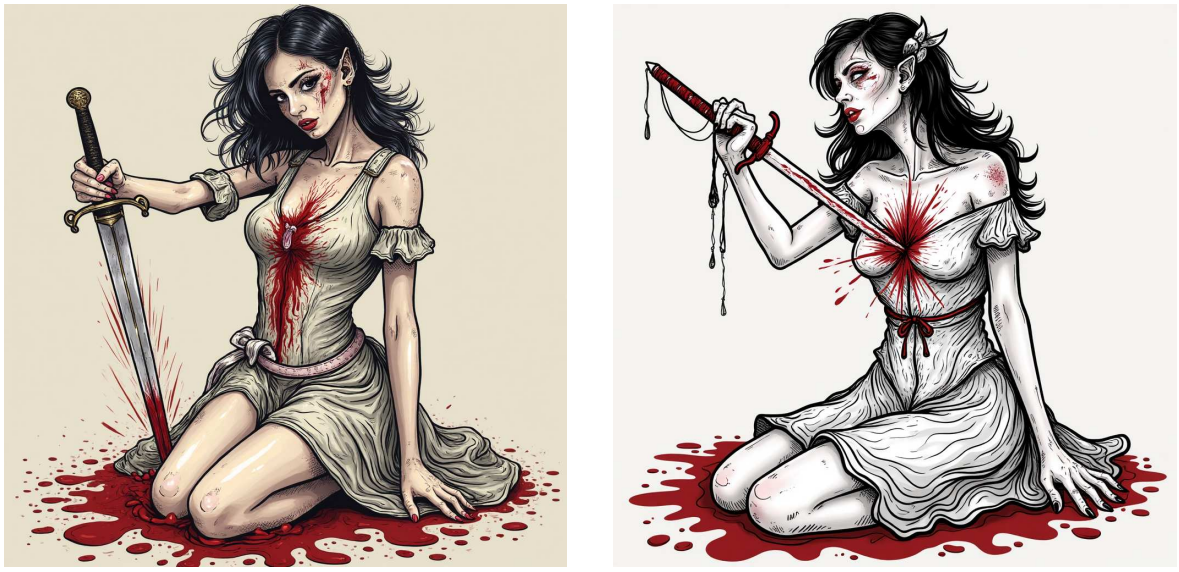
5.1.1 Usability and Prompt-Based Control

P1 did not describe the interface itself as difficult to operate. In the debrief, they said that the system was "pretty easy to use", but "kind of confusing with the prompts". They could run the workflow, but finding prompts that produced a precise visual intention was harder.

The clearest example was their attempt to generate a character stabbing themself in the chest. P1 first described the desired action in ordinary language, but the workflow repeatedly failed to produce the intended pose. During the session they reacted to this as if the system were resisting the request:

> *"This is so weird [...] it's like they're sentient... and they are resisting."*

The desired direction became more reachable only after the prompt shifted, with researcher support, toward a different, more model-effective term, "seppuku". For P1, the difficult part was finding words that would make the model produce the intended visual result. In the later debrief, P1 referred back to this episode as a moment of frustration. When asked about the biggest barrier to using the tool, they identified prompting, while noting that they could not tell whether the difficulty came from their own inexperience or from the system producing unreliable results.

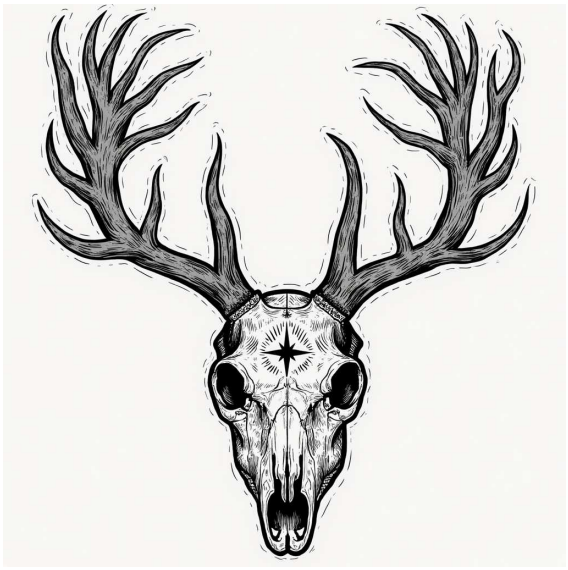


Before the vocabulary shift | After the vocabulary shift

Figure 5.1: Example of a prompt vocabulary shift during P1's session.

A related problem appeared with spatial and geometric details. When P1 prompted for a four-pointed star between the eyes, the workflow often placed star-like shapes in the wrong location or generated different kinds of stars. Some outputs were closer than others, but in this episode the result depended on repeated variation rather than predictable control.

Incorrect star shape



Closer but still imperfect

Figure 5.2: Example of inconsistent spatial interpretation in P1's session.

The controls for the different personalization and image-conditioning paths also required interpretation. P1 described the numerical controls as difficult to connect to expected visual outcomes:

> *"The controls are so random that I feel like you would have to spend a lot more time to understand how to change it to get exactly what you want."*

In one episode, the LoRA-only path produced a more generic realistic result, and the Redux path was used to bring the output closer to the intended style. This recovery depended on knowing which workflow path to try. In the supported session, the researcher could help interpret those options.

Weak style adherence

Stronger style conditioning

Figure 5.3: Using a different conditioning path to recover stronger style adherence.

5.1.2 Creative Utility

P1's strongest positive use case was not finished-image production, but reference generation. They described the appeal as being able to get "an anatomical reference without having to look up something that actually exists already". Later in the session, they explained that in a real drawing process they "would just take the first pose that looks good and recreate it" in their own style. The personalized style made this reference use more acceptable to them: P1 said that they did not like other AI styles and would not use them for references in the same way. They also drew a boundary between reference use and finished-image use:

> *"I'm impressed by the images that already come out looking amazing and not AI generated, but, I don't know, I feel like that defeats the purpose [...] I would prefer to use it for the sketching phase and the references and not the complete image."*

The session also showed a different kind of difficulty from prompt refinement. At several points, P1 did not only need better wording for an existing idea; they did not know what to ask for in the first place. The workflow could respond once P1 gave it a direction, but it did not help much when the problem was getting from "I don't have any ideas" to a usable starting point.

As a side attempt during the session, P1 also tried using ChatGPT for ideation. The suggestions were not satisfying to them: "These suck [...] They're not very creative". They also disliked the tone of the ChatGPT interaction. In this session, ChatGPT did not bridge the gap between having no starting idea and having a useful prompt direction.

Another limit appeared when P1 tried to use an existing sketch as material for further development. They wanted to put their sketch into the workflow and have the system add or develop something around it when they were out of ideas, or in P1's words, to "just put something in there". The available IP-Adapter path could condition generation on the uploaded sketch, but it did not edit the sketch in place. It generated a new stylized image instead. Increasing the conditioning strength did not turn this into sketch-based editing; when pushed further, it produced a low-quality, disfigured result.

Sketch input

Generated output

Figure 5.4: Attempt to use an existing sketch as a basis for further generation.

P1 was also cautious about support that would complete too much of the drawing process. When the researcher described a possible feature where the system would fill in a drawn shape, using the outline of a skeleton as an example, P1 rejected that kind of completion:

> *"I wouldn't like that. I like doing the skeleton myself."*

This kept the desired support closer to references, ideation, and additions to existing material than to automatic completion of the drawing itself.

P1 also described value that did not involve a single specific image. Seeing how the system arranged elements helped them think through spatial relations in their own composition:

> *"I put this here because I was trying to imagine what could be the result here if it was going to turn this into a character [...] When you see the results, you can kind of understand how it places things in relation to each other, and then you can imagine if you had this, like, head here, that it would maybe have, like, this curved spine around it."*

5.1.3 Personalization and Ambivalence

The LoRA personalization was successful from P1's perspective. They rated style matching at 5 out of 5 and recognized specific details from their own drawing habits in the generated images. One example was the way the model rendered the white part of the eye socket, which they identified as something they do in their own work without having explicitly prompted it:

"Oh, it did that thing that I do with the eyes without me telling it to do that [...] I was thinking how I was going to make it do that, but it just happened."

Dataset example Generated output

Figure 5.5: Example of a personalized stylistic detail recognized by P1, shown alongside a related detail from the dataset.

At the same time, some high-fidelity personalized outputs created discomfort. P1 described some outputs as impressive very early in the session, and later explained that the strongest style-matched results could feel "scary" and demotivating when they appeared to do parts of the work better than they felt they could do themself:

"It's scary more so in the way where sometimes it does a better job than you feel like you could have ever done [...] Because it makes you feel like, oh, well, like, why should I even try to do anything?"

This ambivalence was not their whole reaction to the workflow, since they also described clear reference and ideation uses. It was tied especially to outputs that looked polished enough to compete with finished work.

Figure 5.6: Examples of high-fidelity personalized outputs from P1's session.

This ambivalence also appeared in their discussion of professional labor and value. When viewing the outputs in Figure 5.6, P1 acknowledged that an image could be close enough to send to a client with minimal editing: "It looks quite ready to be used". But using it in that way would have been ethically uncomfortable for them. They said that it would feel like cheating:

> *"It still feels a bit like cheating [...] It's just not the same effort at all. So maybe if you charged very little, then you could be doing this."*

For P1, the amount of manual work remaining after generation mattered. Using the system to support a drawing process was acceptable, while using a polished generation as a deliverable raised questions about disclosure, pricing, and the value of manual effort. They connected this directly to the client paying for the artist's time spent drawing: "if they're paying you [...] then they're paying for your time that you spend drawing". They also said that in such a case, the client should know that the image had been generated.

5.1.4 Control, Privacy, and Ownership

The private character of the workflow mattered strongly to P1. They did not describe privacy as an abstract preference only, but as a condition that affected whether they would be willing to use a personalized system trained on their own images. They expressed distrust of corporate data handling:

> *"I wouldn't want to run it on some company's whatever (servers) [...] it would have to be trained on your images and I feel like I just don't trust anyone with that."*

Their position was also pragmatic. They reasoned that in their current situation, a small number of images might simply blend into a large corporate dataset and not matter much. However, they also said that if they were more professionally visible and had thousands of images, they would probably not use such cloud-based tools:

> *"But yeah, if I was very popular and I had thousands of images, then I wouldn't even touch anything like this, probably."*

For this participant, privacy was therefore tied to the perceived risk of exposing their own portfolio material. P1 described that risk as more serious when imagining a larger public portfolio or higher professional visibility.

P1's comments on ownership were positive but not simple. Their first reaction was that they would like the outputs to be theirs, while also treating the question as uncertain. When the researcher later asked whether the prompter or the company should own an image, they said that "the prompter should own the image". At the same time, they recognized that ownership was philosophically uncertain because the image depended on prompts, model training, and source material. They said:

> *"I mean, I would like to own them. Yeah, I think they should be mine, but whether they really are mine is questionable."*

The interview therefore complicates the positive questionnaire score. P1 leaned toward user ownership, especially in a personalized workflow, but did not treat the question as straightforward.

5.1.5 Summary of P1 Findings

P1's session suggests that the original workflow was successful as a supported, personalized source of visual references, but less successful when P1 wanted precise visual control. The interface was learnable enough for the session, and the personalized outputs were recognizable and useful to them. The main barriers were prompt wording, spatial specificity, parameter opacity, and the absence of an editing or transformation path for working from an existing sketch.

P1 did not need the system to replace drawing. They wanted references in their own visual language, help when they did not know what to add, and ways to develop existing material without automating the parts of drawing that they valued. At the same time, the session shows that, for P1, highly polished personalized outputs could also become uncomfortable. Private local deployment made the workflow more acceptable to P1. For this participant, the most acceptable uses were reference generation, sketch-phase support, and possible additions to existing material rather than finished-artwork generation.

5.2 Results: Participant 2 — The 3D Mapping Artist

P2 approached the workflow from the perspective of 3D mapping, animation, and VJ-style practice. They already used generative AI tools in parts of their creative work, including asset generation, animation, compositing, and mapping. In the session, they chose open-ended ideation rather than a fixed brief. P2's questionnaire responses were also positive overall. Their System Usability Scale score was 90, and their project-specific ratings were high for creative utility and personalization. The lower control and trust aggregate mainly reflected a low ownership rating, which is discussed later in the section. As in the P1 case, these scores are used only as descriptive context.

Metric	Score	Descriptive interpretation
Usability (SUS)	90 / 100	Very high perceived usability, with full control still described as something that would require more time.
Creative utility	4.0 / 5.0	Positive perceived usefulness for experimentation and stylistic synthesis, but limited fit for structure-preserving professional tasks.
Personalization	4.5 / 5.0	Strong perceived style match, including recognition of specific visual features from the provided material.
Control and trust	3.0 / 5.0	Mixed aggregate score, shaped by a low ownership rating rather than a simple lack of trust in the local workflow.

Table 5.2: Descriptive questionnaire summary for Participant 2.

5.2.1 Usability and Workflow Control

P2 treated the basic interface as learnable. In the debrief, they described the system as easy to understand at the basic level, while also saying that they did not yet feel they had complete control over it:

> *"Quite simple, but at the same time [...] I don't feel I have everything completely under control."*

When the researcher summarized this as easy to learn but hard to master, P2 agreed. P2 could operate the workflow, but full control over the settings and generation paths would take more time.

P2 hesitated around image navigation controls and strength settings for the personalization and image-conditioning paths. These were not major breakdowns in the session, but they made the workflow feel easy to start and slower to master.

P2 also used ChatGPT during the session to generate prompt sets, then brought those prompts back into the workflow for further image generation. They described this as an existing habit from their broader work with text-based generators:

> *"[...] I rarely write the prompts myself, but usually I have a set of prompts generated somewhere in ChatGPT and I just throw them in there."*

P2 also separated everyday operation from technical maintenance. They indicated that they would not need someone else in order to use the interface, but that training or updating the personalized model was a different matter:

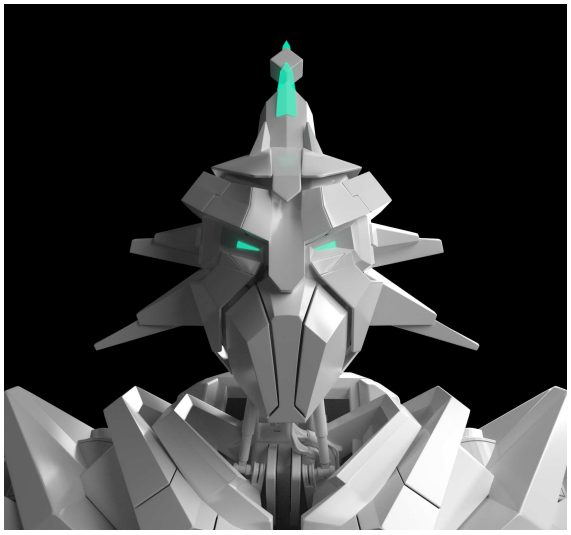
> *"For using it, I would say I don't need anyone, but at the same time, I wouldn't know how to train it (the model) myself."*

For P2, this mattered because their work moves between project-specific aesthetics rather than one fixed style. The supported session could show how the prepared model behaved for one set of reference images, but it did not make training or updating a style model something P2 could independently repeat.

5.2.2 Creative Utility and Structural Limits

A central episode came from P2's mapping practice. They brought a 3D robot render and wanted the workflow to restyle that existing asset while keeping its geometry intact:

> *"I would need to put the robot in there [...] so that it actually preserves the thing and adjusts him [...] to the specific learned style."*



Input render

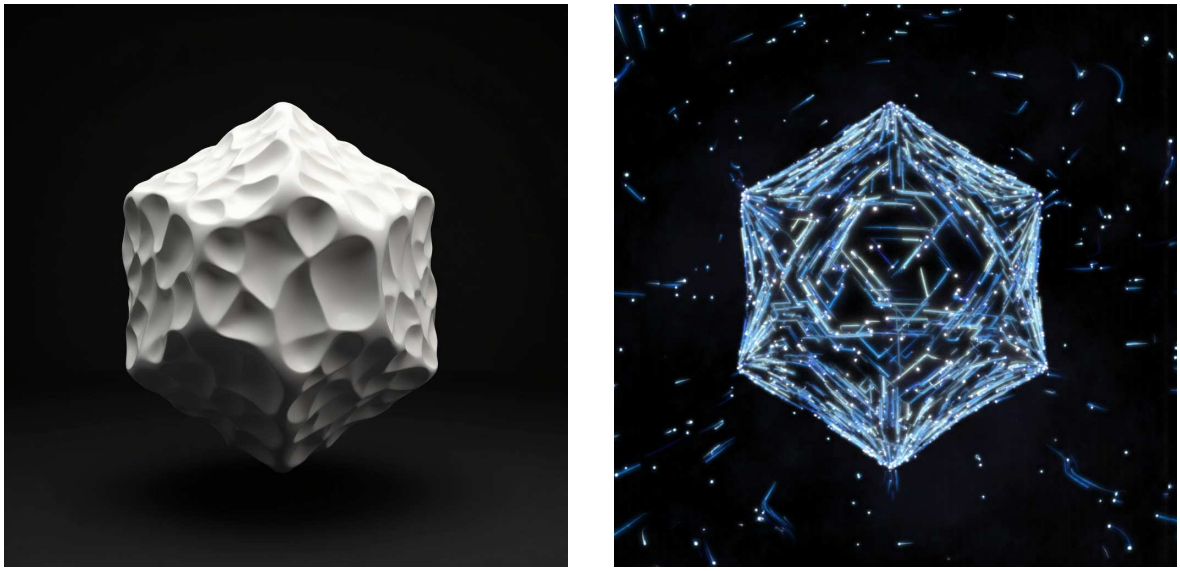
Generated output

Figure 5.7: Attempt to use an existing 3D render as a structure-preserving input.

The available image-conditioning paths did not behave like precise texture application or editing. They mixed visual features from the input with the learned style and produced a new image. This could be visually interesting, but it did not preserve the robot as a

stable object. For P2's practice, this mattered because projection mapping and related work can depend on exact shapes, masks, resolutions, and aspect ratios. P2 described the need to prepare a specific object first and apply a style to it, or to work with a specific shape that the generation would follow. A generated image that looks good may still be difficult to use if it does not fit the object, screen, or spatial layout for which it is being made.

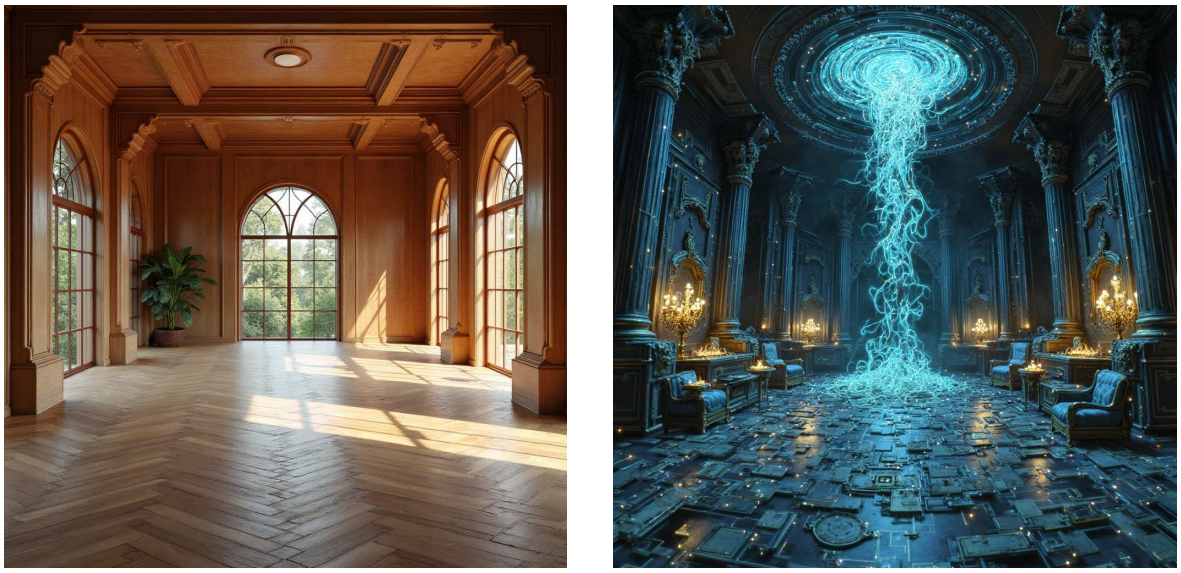
This also shaped P2's use of the workflow controls. At one point, they wanted to see the underlying composition before applying the personalized style. With researcher support, P2 used the node bypass function to first generate a base image and then enable the style path afterward. This did not give them the structure-preserving transformation they wanted, but it created a partial workaround in which structure and style could be explored separately.



Base composition Stylized result

Figure 5.8: P2's staged use of the workflow after the style path was bypassed and re-enabled.

Although the workflow did not solve the robot use case, P2 kept experimenting with how the robot input, the LoRA, and the style-conditioning path affected each other. The result was closer to stylistic synthesis than to controlled transformation, but it produced a visual direction that P2 treated as worth exploring.

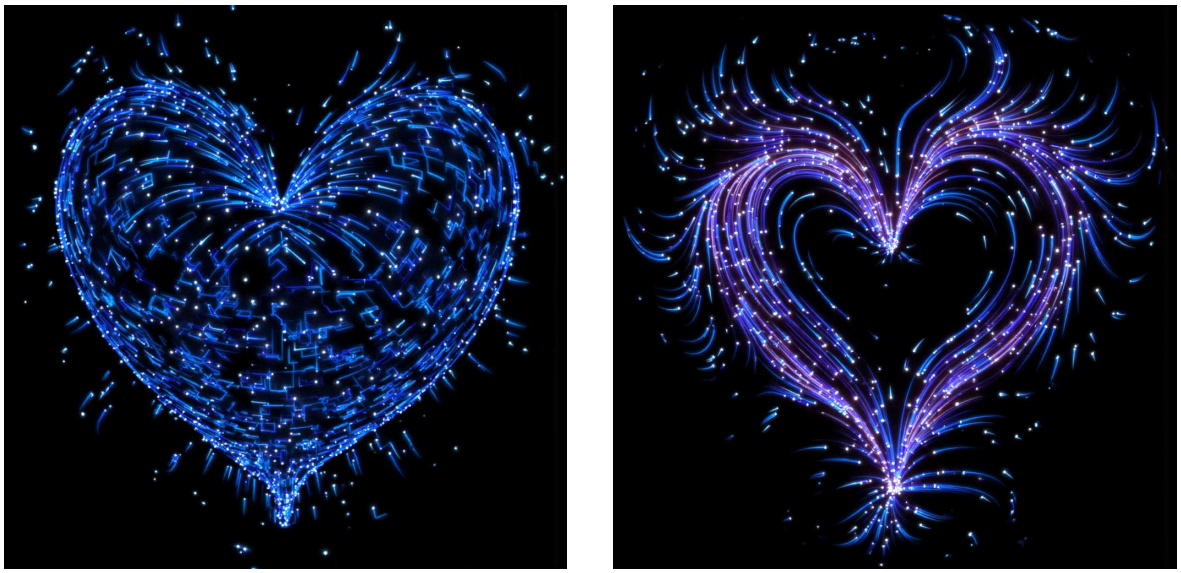


Base generation Stylized synthesis

Figure 5.9: Example of stylistic synthesis produced by combining P2's reference material with the personalized style.

P2's engagement also continued into the debrief. After the formal task phase, they kept testing settings around a heart-shaped image and described wanting to continue interacting with the workflow:

"I still feel like clicking something."



Earlier result Later variation

Figure 5.10: P2's continued exploration of a heart-shaped image across different settings.

This did not mean that the workflow solved P2's professional use case, but it did indicate exploratory value.

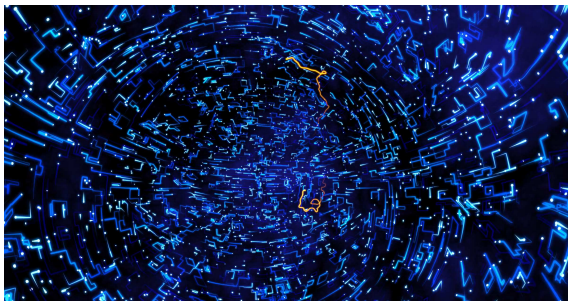
5.2.3 Personalization and Stylistic Recognition

P2 recognized their provided visual style in the generated images. In the debrief, they said:

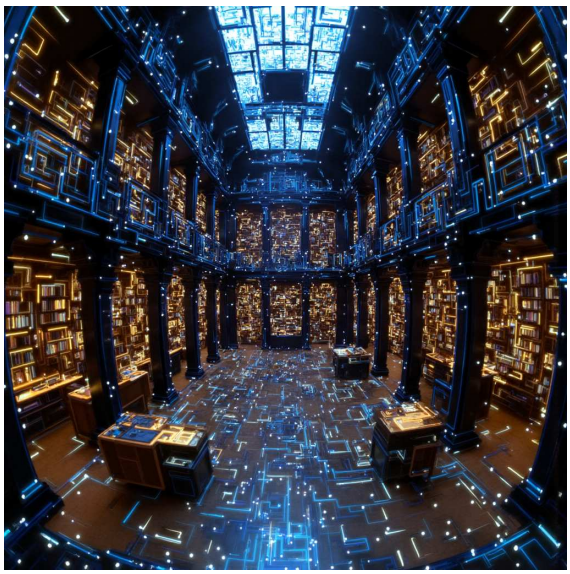
> *"I see in them the style I gave you [...] the style is definitely there."*

This recognition was also visible during the session. In one episode, P2 noticed that the workflow had carried over orange elements from a reference image into an output, even though the broader visual style was dominated by cooler blue tones. They reacted to this as a specific and somewhat surprising stylistic pickup:

> *"That's funny that it took from those – those are two beings, those orange things – and it actually took from them, that other color."*



Reference image



Generated output

Figure 5.11: Example of a smaller color feature from P2's material appearing in a generated output.

The personalization was also meaningful to P2 because it differed from their previous experience with generative AI tools. They described the session as the first time they had encountered an AI model trained on their own style, rather than selecting a style trained by someone else:

> *"It is definitely interesting in that I never really had an AI trained on my own style. I actually encountered that for the first time."*

For P2, the personalized model therefore added a kind of specificity that more generic style presets had not provided.

5.2.4 Control, Privacy, and Ownership

P2's comments on privacy were pragmatic and tied to professional material that could not be shared publicly:

> *"[...] most of the time I don't care, but sometimes it's necessary, when I can't publish certain things."*

They also expressed interest in running the workflow at home:

> *"[...] I quite enjoy it, I would like to be able to run this at home."*

The questionnaire score for ownership was low, but the interview suggests that this was not simply because P2 felt a lack of agency. P2 described a broader stance toward AI creation in which they did not strongly claim ownership over these kinds of outputs:

> *"I don't perceive ownership in this [...] I perceive it all more so in a kind of open source spirit."*

This helps explain why the control and trust aggregate was lower than the rest of the questionnaire profile. For P2, not wanting to claim ownership did not necessarily mean that the workflow felt untrustworthy. It reflected their broader stance toward generated material.

5.2.5 Summary of P2 Findings

P2's session suggests that the original workflow was engaging and learnable as a supported personalized generation system, especially for experimentation with style. P2 recognized their provided visual style in the outputs, noticed specific features being carried through, and continued to explore settings after the formal task phase had ended. At the same time, the session also points to a clear limit for their professional use case. P2 needed a way to apply style to an existing object while preserving its geometry, but the available conditioning paths produced new syntheses rather than structure-preserving transformations.

P2 could operate the prepared interface, but parameter interactions, graph-level control, and especially training or updating new styles remained outside what they felt able to do independently. Local deployment mattered mainly as operational flexibility and confidentiality for client-sensitive work. Ownership was less central to their acceptance of the workflow than the ability to experiment with a personalized style under their own control.

5.3 Cross-Case Synthesis

The two sessions do not point to one single use case. P1 used the workflow mainly for reference generation, sketch-phase support, and moments of uncertainty about what to prompt. P2 used it for open-ended exploration and stylistic experimentation, while also testing whether it could restyle an existing 3D object for use in a project pipeline. In both cases, the workflow was engaging enough to support continued exploration in the session. The useful outputs were often material to react to, adapt, or carry into another process rather than finished results. Its limits became visible when participants needed it to preserve a specific intention, structure, or part of their process.

5.3.1 Usability and Control

Both participants described the workflow as relatively easy to begin using, which fits the high SUS scores. The interviews and session episodes still show a difference between perceived usability and control. P1 could use the interface, but struggled to find prompt wording that produced precise visual relations, actions, and spatial details. P2 also understood the basic workflow, but did not feel that they had full control over the interaction between input images, personalization strength, and the different generation paths.

Starting the workflow and steering it reliably were not the same thing. The workflow had a low enough entry point for both participants to explore it in the supported session, but its behavior was still difficult to predict. Some of this difficulty came from exposed parameters whose visual effects were not immediately clear. Strength values for the LoRA, Redux, and image-conditioning paths gave participants control in principle, but the meaning of a small or large adjustment had to be learned through trial and error. Some of the difficulty also came from the models themselves, which could produce good images while still missing the specific constraint that mattered to the participant. So the problem was not only the interface. It was also hard to predict how prompts, references, and settings would combine.

5.3.2 Prompting and Semantic Support

The two participants used language support differently. For P1, prompting was often a barrier. They sometimes knew the kind of image they wanted, but had to search for wording the model responded to, and the model did not reliably follow precise visual constraints. The self-stabbing and four-pointed-star episodes show this problem clearly. Ordinary language could describe the intention, but it did not reliably produce the image.

P1 also tried using ChatGPT for ideation when they were unsure what kind of image

to make next. This did not give them a useful direction to bring back into the image workflow. The suggestions felt generic and stylistically mismatched. The difficulty was not just that P1 lacked language support. In this episode, generic ideation support did not translate into a useful next step for the image workflow.

P2's relation to language support was different. They already used ChatGPT in their wider practice to generate prompt sets for text-to-image tools, and they repeated this pattern during the session. For P2, LLM-assisted prompting was already part of how they worked with generative image systems. In the evaluated workflow, however, that support sat outside the interface. P2 had to leave the generation workflow, produce prompt material elsewhere, and bring it back in.

5.3.3 Generation Versus Transformation

Both sessions also exposed a limit in a workflow built mainly around generating new personalized images. Both participants found value in generation, but both also moved toward tasks where they wanted to change or extend existing visual material.

For P1, the desired use was close to sketch-phase support. They wanted references in their own visual language, but they also wanted to put an existing sketch into the workflow and have the system help add or develop something without replacing the drawing process. They wanted support around parts of the process where they were uncertain, while preserving the parts of drawing that still mattered to them.

For P2, transformation meant something different. Their use case was closer to restyling, texturing, and pipeline integration. They needed a generated result that could fit a specific object or display context. A good-looking blended image could still be hard to use professionally if it did not preserve the structure required by the later mapping or compositing stage.

The details differed across the two cases. For P1, the structure to preserve could be a pose, a spatial relation, an existing sketch, or a part of the drawing process. For P2, it could be the geometry of a 3D render, a projection surface, an aspect ratio, or a mask. In both cases, visual quality alone was not enough. A generation-first system could not easily work on existing visual material while keeping the constraints that made that material useful.

5.3.4 Personalization and Relevance

Both participants recognized the personalized style in the outputs. P1 rated the style match very highly and noticed specific drawing habits reproduced by the model. P2 also recognized their supplied style and described the experience as novel compared with using styles trained by someone else. In both cases, personalization made the

workflow feel less generic because the outputs reflected material the participants had supplied and could recognize.

This mattered in different ways. For P1, personalization made generated references more acceptable because they were closer to their own visual language than generic AI styles. For P2, it made the workflow engaging and distinctive because it offered something beyond selecting an existing style preset. The limit was that this recognition did not settle the usefulness of the output by itself. Some polished outputs caused mixed feelings for P1, and P2 still needed results that could fit their pipeline.

5.3.5 Privacy and Ownership

The private setup mattered in both cases, but not for the same reason. For P1, privacy was mainly protective. The concern was that a personalized system trained on their own images would expose portfolio material to a company or a large external dataset. Local deployment made the workflow more acceptable because it reduced that perceived risk.

For P2, privacy was more operational. They were not strongly concerned about privacy in all creative experiments, but local or offline use mattered when client material could not be shared publicly. In this case, the value was less about protecting an artist's portfolio or style from being used in someone else's system, and more about being able to use the tool inside professional constraints.

The ownership responses also diverged. P1 leaned toward wanting the outputs to belong to the user, while still recognizing that the question was uncertain. P2 did not strongly claim ownership over generated images and framed this as part of a broader open-source orientation.

5.4 Limitations and Lessons Learned

The methodological limits of the first evaluation were discussed in Chapter 3, so they are only summarized here in relation to the findings. The results come from two purposively recruited participants using a prepared workflow in researcher-supported sessions. They do not show how artists would install, maintain, or personalize the system independently, or whether the same workflow would be practical on different hardware setups. They should not be read as describing artists in general. The questionnaire scores are useful descriptive context, but the main evidence is the qualitative material from the sessions, debrief interviews, and generated images. The interpretation is also shaped by the researcher's role in building, supporting, and analyzing the workflow, and by P1's prior familiarity with the project.

Within that scope, the first workflow worked for a narrower kind of use than the thesis

aims to support. It could produce recognizable stylistic outputs and support exploratory work in a setting with researcher support. It was less useful when participants wanted to edit existing material, preserve a specific structure, or understand how prompt and parameter changes would affect the output.

The main lesson for the redesign is that personalized generation and transformation need to be treated as different kinds of support. Personalized generation can remain useful where the artist wants references, variations, or possible directions. Editing, restyling, and transformation are needed where the artist wants to preserve a sketch, object, pose, mask, or composition. The redesign should also reduce the burden of interpreting prompts and parameters, while keeping the local/private character that made the workflow more acceptable in both cases.

Chapter 6

Redesigned Workflow: From Generation to Transformation

6.1 Rationale for the Redesign

The first evaluation suggests that the original workflow was useful as a supported personalized generation system, but too narrow for some of the work participants tried to do with it. It could produce recognizable stylized outputs and support exploratory reference generation. It was less suited to cases where the participant wanted to change existing material while preserving a sketch, object, pose, or composition.

The redesign therefore did not replace generation with editing. It kept personalized generation as one part of the workflow, because Chapter 5 showed that generated references, variations, and unexpected visual directions could still be useful. The aim was to extend that workflow so that generation and transformation could be handled in the same local environment. This meant adding prompt-based editing, reducing the number of exposed conditioning paths, integrating local prompt mediation, and preserving the private deployment that supported the workflow's value. The main changes are summarized in Table 6.1.

Design need	Implemented change
Keep personalized generation while reducing workflow complexity.	Dedicated generation mode using one FLUX.2 [klein] model and participant-specific LoRA, with LoRA strength as the main exposed personalization control.
Add transformation support.	Prompt-based editing mode for working from generated, uploaded, or previously saved images.
Make generated outputs usable as material for further work.	Image history and upload paths for selecting generated, saved, or external images as editing inputs.
Reduce prompt-writing burden.	Local LLM mediation for generation and editing prompts.
Support less finished exploration.	Finished and work-in-progress generation modes.
Preserve local/private use.	Local image model, LLM, and LoRA training.

Table 6.1: Main changes introduced in the redesigned workflow.

6.2 Overview of the Redesigned Workflow

The redesigned workflow was implemented as a new ComfyUI graph with two main modes: generation and editing. The interface keeps these modes visually separated. The generation path is placed on the left side of the visible workspace, while the editing path is placed on the right. A mode switch is used to turn the relevant group of nodes on or off, so the user can move between creating a new image and modifying an existing one without loading a separate workflow.

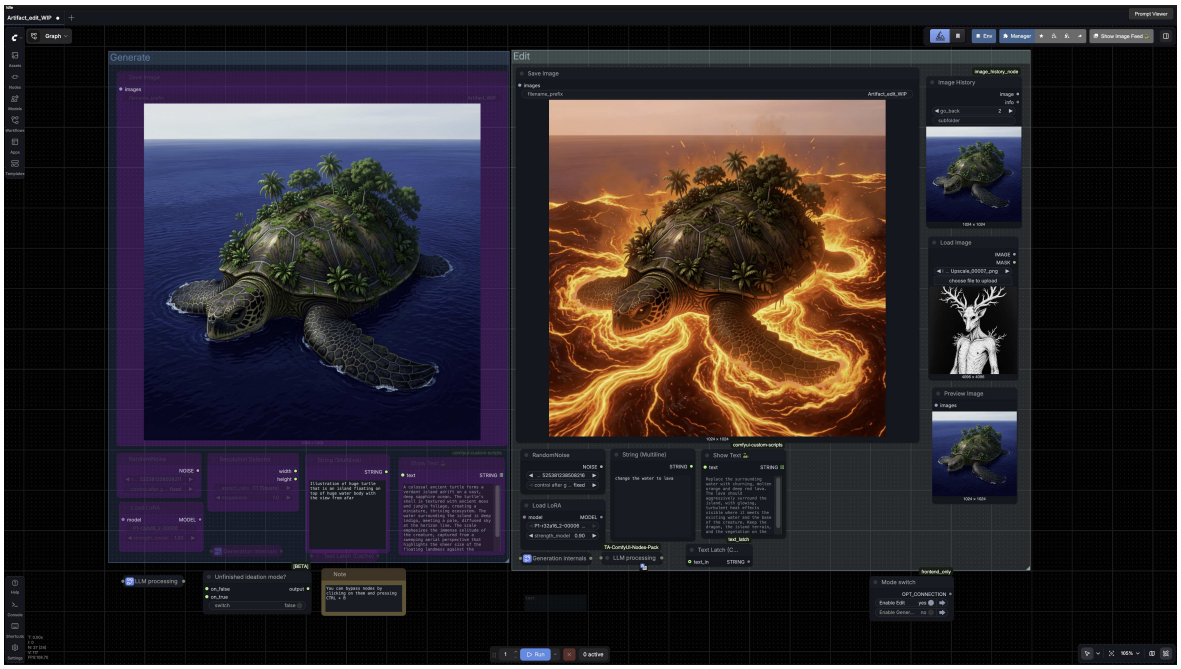


Figure 6.1: ComfyUI interface for the redesigned workflow.

After the first workflow had been implemented, FLUX.2 [klein] became available. Its documentation presents it as a fast, open-weight FLUX.2 variant with support for both text-to-image generation and multi-reference editing (Black Forest Labs, 2026b,c). This made it a practical candidate for the redesign because one of the main limits of the first workflow was already clear: participants did not only want to make new images from prompts or references, but also to work on existing images. The redesigned workflow therefore uses the official FP8 version of FLUX.2 [klein] 9B-KV. The model is used in both generation and editing mode, while the more detailed limits of prompt-based editing are discussed later in the chapter.

Stylistic personalization remains part of both modes. A participant-specific LoRA can be loaded into the FLUX.2 model, and LoRA strength remains the main exposed personalization setting. Redux and IP-Adapter were not carried over into this version, so LoRA is the main explicit style-control mechanism in the redesigned interface. The updated LoRA training pipeline and the reasons for removing the earlier reference-conditioning paths are described in Section 6.6.

The workflow also integrates a local vision-language model as a prompt mediator. The mediator can rewrite prompts before generation or editing, while still allowing direct prompting when mediation is not wanted. Because generation and editing require different kinds of instructions, their prompt-mediation designs are discussed separately in Sections 6.3 and 6.4.

The two modes are connected through saved images rather than through a direct loop in the node graph. Generated and edited images are saved to the ComfyUI output directory, and the editing mode can load recent outputs or uploaded external images as its input. This file-based handoff makes iterative editing possible while keeping generation and editing as separate node groups. The custom nodes and preview tools that support this interaction are described in Section 6.5.

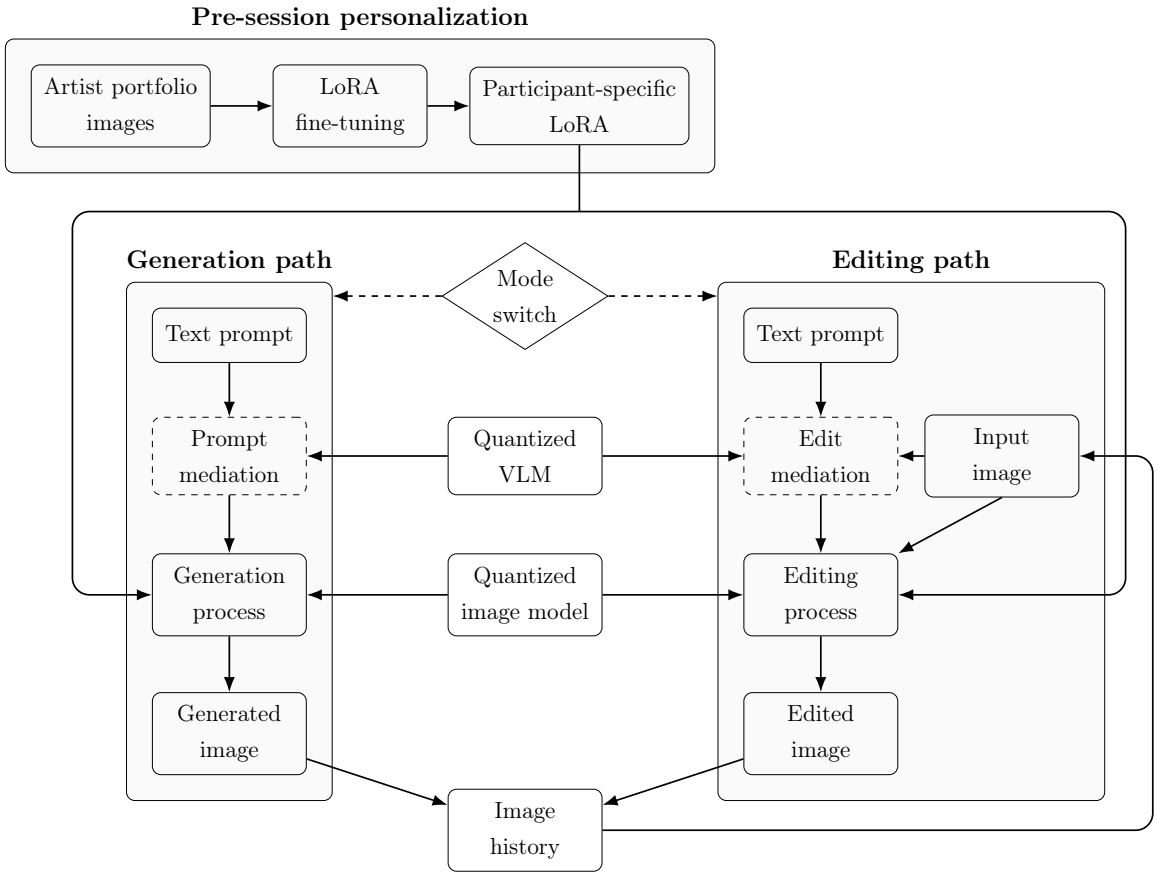


Figure 6.2: Simplified structure of the redesigned workflow. Dashed boxes indicate optional semantic mediation.

This architecture should still be read as a research workflow rather than as a finished artist-facing application. The participant interacts with the ComfyUI graph through a researcher-arranged interface, and the workflow still depends on technical setup, model files, LoRA training, and local server configuration. The purpose of the redesign is to prepare an artifact for testing whether an integrated generation-and-editing workflow can support kinds of interaction that were not possible in the first evaluation.

6.3 Generation Mode

Generation mode keeps the basic role of the first workflow: it produces new images from a text prompt using a participant-specific LoRA when personalization is wanted. In the redesigned workflow this path uses the same FLUX.2 [klein] 9B-KV FP8 model as the editing path, rather than a separate generation-only model. The main user-facing controls are the text prompt, the LoRA strength, and the output aspect ratio. LoRA strength controls how strongly the participant-specific style adapter is applied, while the aspect-ratio setting allows the user to prepare images for different formats without changing the prompt text itself.

The prompt can be used in two ways. In direct prompting mode, the user's text is passed to the image model without LLM rewriting. In mediated mode, the prompt first goes through a local Gemma 4 E2B vision-language model running as a prompt mediator. The mediator is not used as a conversational agent here. Its task is narrower: it rewrites the user's prompt into a fuller FLUX.2 prompt while keeping the prompt inspectable before image generation. The generation prompt rules were adapted from the official FLUX.2 [klein] prompting guide, which recommends descriptive natural-language prompts rather than short tag lists (Black Forest Labs, 2026e). In the workflow, this means that a short request can be expanded with subject, setting, visual detail, lighting, and atmosphere, but the user can still bypass the mediator if the rewritten prompt is not useful.

The generation side also includes two prompt-mediation modes: finished and work-in-progress. Finished mode is the default mode and asks for a complete, polished image. Work-in-progress mode asks for an image that appears unfinished, partially rendered, or still being developed. This option was added for two reasons. First, P1's reaction to some highly polished personalized outputs showed that impressive finished images can be ambivalent rather than simply useful. Second, the discussion of fixation in Chapter 2 suggests that polished examples can become anchors for later work. The work-in-progress mode is therefore a design response to these concerns, not an evaluated solution to them.

The work-in-progress prompt rules were kept deliberately narrow. Earlier prompt versions tended to turn the image into a concept sheet or studio document, adding paper margins, annotations, generated handwriting, tools, or surrounding process material. The current rules instead ask for the subject itself to look unfinished: for example through rough blocking, visible construction lines, incomplete shading, exposed underpainting, flat colour passes, loose edges, or partially refined details. This keeps the mode closer to the intended use as provisional visual material, rather than turning it into a separate graphic-design format.

6.4 Editing Mode

Editing mode is the right-side transformation path of the redesigned workflow. It uses FLUX.2 prompt-based editing rather than a separate inpainting or mask-based system. The input image can be the latest generated image, a previous image selected from the output history, a previous edited result, or an uploaded external image. This makes the editing path usable both after generation and as a standalone way to work on existing visual material.

The editing prompt can also be mediated by the local vision-language model. In

this case, the model receives the selected image together with the user's short edit prompt, and rewrites it. The editing mediator uses a shorter instruction set than the generation mediator because the image model already receives the source image. Instead of describing a whole scene, the rewritten prompt is meant to state what should change and, where possible, what should remain unchanged. For example, a short instruction such as changing a colour, adding an object, or adjusting a feature can be rewritten with explicit preservation wording about pose, composition, background, or style.

This preservation wording should be understood as an implementation heuristic. During informal development testing, simple edit prompts sometimes changed parts of the image that were not meant to be altered. Prompts that also named what should remain stable tended to produce more consistent edits, although not in every case. For this reason, preservation wording was added to the mediation instructions, while direct prompting remains available when the mediated prompt is not useful. In practice, prompt-based editing can sometimes preserve the image geometry or change a specific part of the image quite consistently. Its control is still different from a mask, depth map, ControlNet-style condition, or dedicated compositing workflow, where the user supplies an explicit spatial constraint. The editing mode therefore addresses part of the transformation gap identified in Chapter 5, but the user specifies the change semantically rather than by marking the exact region or structure to preserve.

LoRA personalization is available in editing mode as well as in generation mode. This allows the same participant-specific style adapter to be used when creating a new image and when modifying an existing one. In practical terms, this keeps personalization attached to the whole generation-and-editing workflow rather than limiting it to the first output. The effect of LoRA strength during editing still has to be evaluated with users, since a stronger style influence may also change parts of the image that the user intended to preserve.

6.5 Interaction Support and Custom Workflow Nodes

Several small support nodes were added around the two main modes. These nodes do not change the image model itself. Their purpose is to make the workflow easier to operate, especially when images and prompts move between the two sides of the ComfyUI workspace. They were implemented as custom nodes because suitable community nodes for these specific history, prompt-latching, and metadata inspection functions were not found during development.

The most important support node is the image history selector. ComfyUI workflows are directed graphs, so an edited output cannot simply feed back into its own input as a graph cycle. The redesigned workflow therefore connects generation and editing

through saved files. The custom Image History node reads from the ComfyUI output directory and exposes a `go_back` index. An index of `0` selects the latest image, while larger values step back through earlier outputs. Because edited outputs are saved into the same output history, the latest image can be a newly generated image or the result of a previous edit. This gives the editing mode a simple way to continue from the most recent result without turning the graph itself into a loop.

Standard ComfyUI Preview Image nodes are also used in the editing path. The Image History node includes its own preview of the selected image, but that preview is updated when the node is interacted with. In some cases the selected index can still point to a different file after a new image has been generated or edited, because the output history has moved forward. For example, an index that previously selected the third most recent image may select a different image after another output is saved. The separate preview node shows the image that is actually sent into the editing model when the workflow is run. This makes the current editing input visible even when the history-node preview has become outdated.

The workflow also includes a Text Latch node for prompt handling. Prompt mediation can produce a useful rewritten prompt, but ComfyUI bypassing and muting can interrupt the normal text path through the graph. The latch stores the most recent prompt text and keeps it available even when both the prompt and VLM mediation nodes are bypassed. This allows a useful mediated prompt to be retained for later generations or edits.

A separate Prompt Viewer was built to inspect previous outputs and their prompt metadata. ComfyUI stores workflow and prompt information inside generated PNG files, but ordinary history views were not flexible enough for this workflow. The redesigned graph contains multiple prompt fields, optional mediated and direct paths, active and bypassed nodes, and both generation and editing outputs. The Prompt Viewer reads the metadata from generated images and exposes selectable text fields, so the user can inspect which prompts were actually used for a given image. This was mainly an interaction-support and documentation tool for the research workflow, rather than a validated user-facing feature.

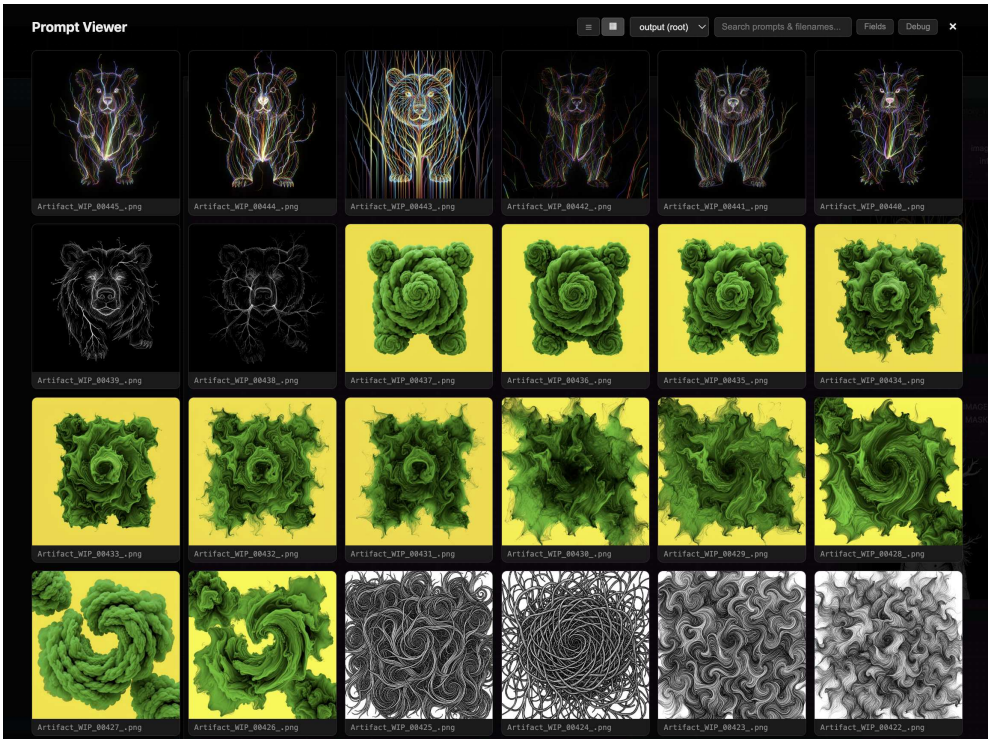


Figure 6.3: Prompt Viewer overview showing saved generated images.

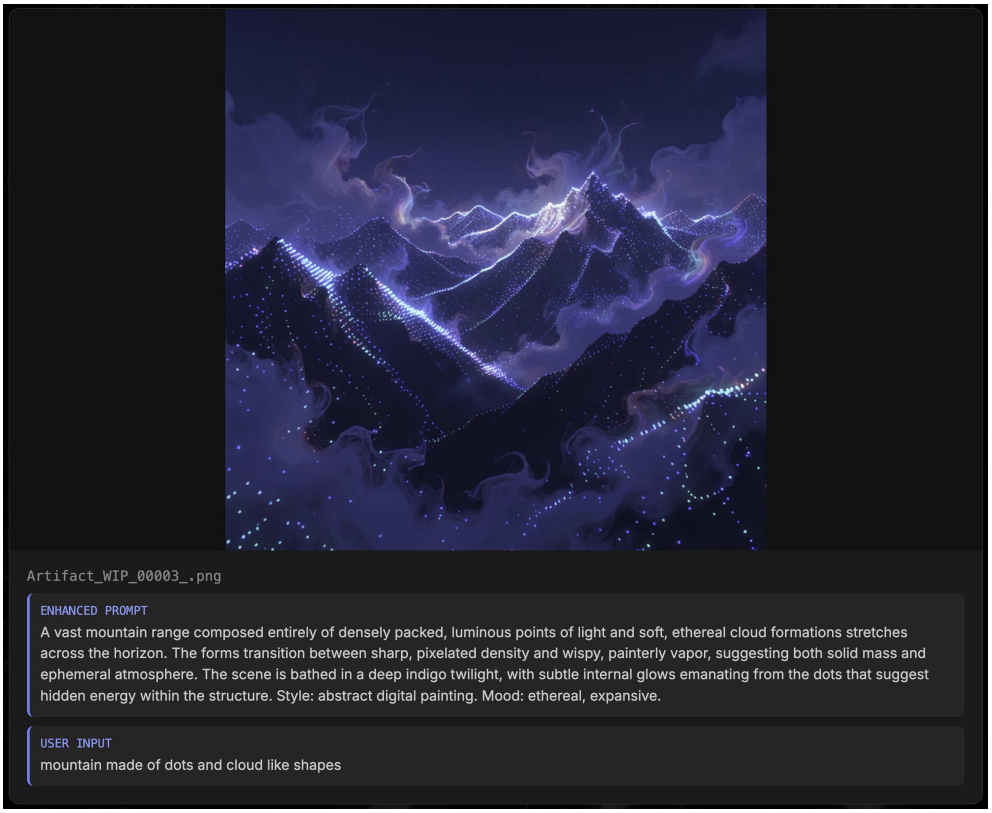


Figure 6.4: Prompt Viewer detail view showing a selected image with its original and enhanced prompts.

6.6 Personalization Pipeline Update

The personalization pipeline was updated because the image model changed. In the first workflow, participant-specific LoRAs were trained with Kohya-SS sd-scripts (Kohya-SS, 2026b). The redesigned workflow instead uses Musubi Tuner, a newer Kohya-SS repository that includes LoRA training support for FLUX.2 [klein] (Kohya-SS, 2026a). The purpose of this change was practical rather than conceptual: LoRA remains the main personalization mechanism, but the training implementation had to match the FLUX.2 model used in the new generation and editing paths.

The rest of the personalization logic is intentionally close to the first workflow. Participant portfolio images are still prepared before the session and used to train a participant-specific style adapter. The image-captioning approach also remained similar, captions were produced with local vision-language-model support and then treated as practical training metadata, rather than as a separate evaluated part of the method. The resulting LoRA can then be loaded into the workflow and used in both generation mode and editing mode. LoRA strength remains the main explicit personalization control exposed in the interface. This keeps the style-control surface smaller than in the first workflow, where the LoRA-only, Redux, and IP-Adapter paths each had their own settings and could each be combined with the participant-specific LoRA.

Redux and IP-Adapter were not carried over into the redesigned workflow. This does not mean that they were unhelpful in the first version. They provided alternative generation routes and reference-conditioning options, and in the first evaluation they were sometimes useful when the LoRA-only path did not apply the participant's style strongly enough. They were removed here for three related reasons. First, the available implementations belonged to the earlier FLUX.1-based setup rather than the FLUX.2 [klein] workflow. Second, keeping three separate generation paths would preserve much of the interface and parameter complexity that the redesign was meant to reduce. Third, the redesigned workflow already combines prompt-based editing with participant-specific LoRA, so some of the earlier reference-conditioning use cases could be approached through a simpler generation-and-editing path.

Prompt-based editing can also be used after an image has been generated to try to increase or correct stylistic influence, for example when an initial generation does not apply the participant's style strongly enough. During informal development testing, the FLUX.2 [klein] LoRAs trained with Musubi Tuner also appeared to apply style more consistently than the earlier implementation. These observations supported a simpler LoRA-centered control scheme for the implemented artifact, but they do not by themselves establish that the updated personalization pipeline works better for participants.

6.7 Implementation Environment and Constraints

The redesigned workflow was implemented as a local ComfyUI setup. The image model is the official FLUX.2 [klein] 9B-KV FP8 checkpoint (Black Forest Labs, 2026b). Prompt mediation is handled by a Q4-quantized Gemma 4 E2B checkpoint, served locally through `llama.cpp` (ggml.org, 2026; Unsloth, 2026a). ComfyUI calls this local server through a community LLM node using the server's OpenAI-compatible API.

The main implementation constraint was GPU memory. The workflow addressed this by using quantized versions of both main models: FP8 for FLUX.2 [klein] and Q4 for Gemma. This made the redesigned workflow feasible on the same RTX 4080 Super with 16 GB of VRAM used for the first implementation, including local inference, captioning, and LoRA training. Because the setup remained local, prompts, input images, LoRA files, generated outputs, and edited outputs stayed inside the researcher-controlled environment.

This remains a researcher-supported workflow rather than an independent consumer installation. Participants use the prepared ComfyUI graph; they are not expected to install dependencies, download checkpoints, configure GPU inference, run `llama.cpp`, maintain the graph, or train LoRA modules themselves. This boundary matters for how the evaluation should be read: it concerns supported use of a prepared local workflow, not independent installation or maintenance of the full technical stack.

Chapter 7

Evaluation of the Redesigned Workflow

This chapter presents the evaluation of the redesigned workflow. The findings are organized first around P1's repeat evaluation and then around P3's session with the redesigned workflow, before turning to a cross-case synthesis. P1's case allows comparison with the original workflow, while P3's case provides an additional perspective on the redesigned artifact. As in Chapter 5, questionnaire responses are used descriptively and the main analysis is based on qualitative session material.

7.1 Results: Participant 1 Re-Evaluation

P1's questionnaire responses after the redesigned workflow session were very positive. Their System Usability Scale score was 90 out of 100. They also gave high ratings for creative utility, personalization, and control, privacy, and trust. As in Chapter 5, these scores are used descriptively. The interview material is needed to interpret what the ratings meant in practice, because P1 also identified specific limits and remaining frictions during the session.

Metric	**Score**	**Descriptive interpretation**
Usability (SUS)	90 / 100	Very high perceived usability, with remaining practical frictions around image history, small text, and technical labels.
Creative utility	4.8 / 5.0	Very high perceived usefulness, especially for brainstorming, art block, and reference generation.
Personalization	5.0 / 5.0	Very strong perceived style match and preference for personalized over generic images.
Control, privacy, and trust	5.0 / 5.0	Very positive ratings for control, local/private use, ownership, and comfort with confidential work, with ownership still discussed in relation to style and source material.

Table 7.1: Descriptive questionnaire summary for P1's redesigned workflow session.

7.1.1 Usability, Prompt Mediation, and Workflow Control

P1 described the redesigned workflow as easy to use overall, which was consistent with the high SUS score. At the same time, some interface elements were hard to see in the session setup, and P1's ease of use partly reflected prior experience with the first workflow and researcher explanation during the session.

The most important usability change concerned prompting. In the first session, prompting was P1's main barrier to control. The redesigned workflow did not remove prompting from the process, but local prompt mediation made it less obstructive. P1 compared the second session with the first session directly:

> *"I remember I was struggling with the prompts before and now it just does it for you kind of and it seems to work pretty well."*

This suggests that prompting became easier to work with in this session. P1 could enter shorter or less model-specific instructions and let the mediator expand them. They also liked that they could write "some like one sentence" instead of having to construct a full image-model prompt manually. Prompting was therefore less obstructive than in the first session, although later episodes showed that it was not fully solved.

In one part of the session, P1 returned to the self-stabbing pose that had been difficult in the first session. In the first session, P1 had to keep changing the text prompt when a result was close in some ways but wrong in others. In the redesigned workflow, P1 generated an image they liked in some respects, but it did not match the intended action: the figure appeared already stabbed rather than actively stabbing themselves, and the gender was also wrong. P1 then used editing mode to try different changes and return to already edited versions when later attempts moved away from the intended result. Later in the editing sequence, P1 also returned to the more model-effective "seppuku" wording that had produced the closest result in the first session, and this led to the final version of the pose that P1 liked most. The difficult part was still getting the image to do exactly what P1 wanted, especially in the pose and action. However, the iterative editing sequence still led to an image that P1 considered useful, though not perfect. Editing gave P1 a way to continue from an image that already had promising qualities. P1 contrasted this with prompt-only regeneration, where changing the prompt could produce either an unrelated image or a nearly unchanged one. Editing made it possible to steer a specific output, while text-only regeneration had often lost the useful parts of the previous image. In the later debrief, P1 described editing in these terms:

> *"No, I think it's quite useful. I'm surprised that it works that well because I was able to get the final image from something that was basically completely different at first."*

Initial generated image | Before vocabulary change | Candidate reference

Figure 7.1: Examples from pose editing episode.

The new features were usable during the session, but they also added more workflow state to keep track of. During the session, P1 moved between generation mode and editing mode, used mediated prompts, and experimented with finished and work-in-progress generation. These functions did not make the workflow difficult to use overall, but they often needed explanation in the moment. Work-in-progress mode was initially understood as if it might be a separate image-level mode, while in the implemented workflow it was a prompt-processing option. The prompt-processing node was explained during the session and could be bypassed, but its small size and placement sometimes made the relevant control hard to locate quickly.

Prompt mediation also left one familiar issue in place. Style wording still mattered. If a prompt led the image model toward generic realism, P1's LoRA style became less visible. The mediator did not remove the need to understand this interaction. In practice, prompt mediation and better prompt wording are hard to separate in this session, because mediated prompts were the main way P1 interacted with the generation and editing paths.

The remaining usability issues were practical. The main barrier P1 identified was the image-history behavior, because they had to repeatedly move back through the history when they wanted to keep editing the same image. New outputs changed the image index after an edit, so P1 sometimes had to move back through the history to continue from the intended image. This did not prevent use of the editing mode, but it made the workflow state harder to follow.

The interface language also stood out. P1 suggested that node labels should be more artist-facing, for example replacing technical terms such as LoRA with labels closer to "your model", "your prompt", or "AI generated prompts". After the session they would remember what to do, but the labels were still closer to ComfyUI terminology than to

the language P1 expected as an artist using the workflow.

The usability result should therefore be read together with the session context. Compared with P1's first session, prompting no longer appeared to be the main obstacle. The redesigned workflow was still a prototype in a supported research setting, and image-history behavior, small text, and technical labels remained minor practical frictions. P1's overall usability judgment was still positive.

7.1.2 Creative Utility Across Generation and Editing

P1 again treated the workflow mainly as a source of reference and ideation material rather than as a way to produce final artwork. When asked where they would use the tool in their process, they said they would mainly use it for brainstorming, art block, and references.

The generation side of the workflow therefore retained the main positive role it had in the first session. P1 could look at an output and judge whether it would be useful as a reference without necessarily drawing from it during the recorded session. In one sequence, P1 generated a mushroom-woman image and focused on details such as hair, face, clothing, and overall pose. They did not treat the output as finished work, but as material that could be changed, redrawn, or combined with their own ideas. In another sequence, an unexpected mask image was not exactly what P1 had intended, but it gave them a new visual idea. In the debrief, P1 said there were more useful unexpected outputs than in the first session. They tentatively connected this to prompt mediation, saying that the prompt support seemed better at formulating model-effective prompts than they were.

Mushroom woman reference candidate

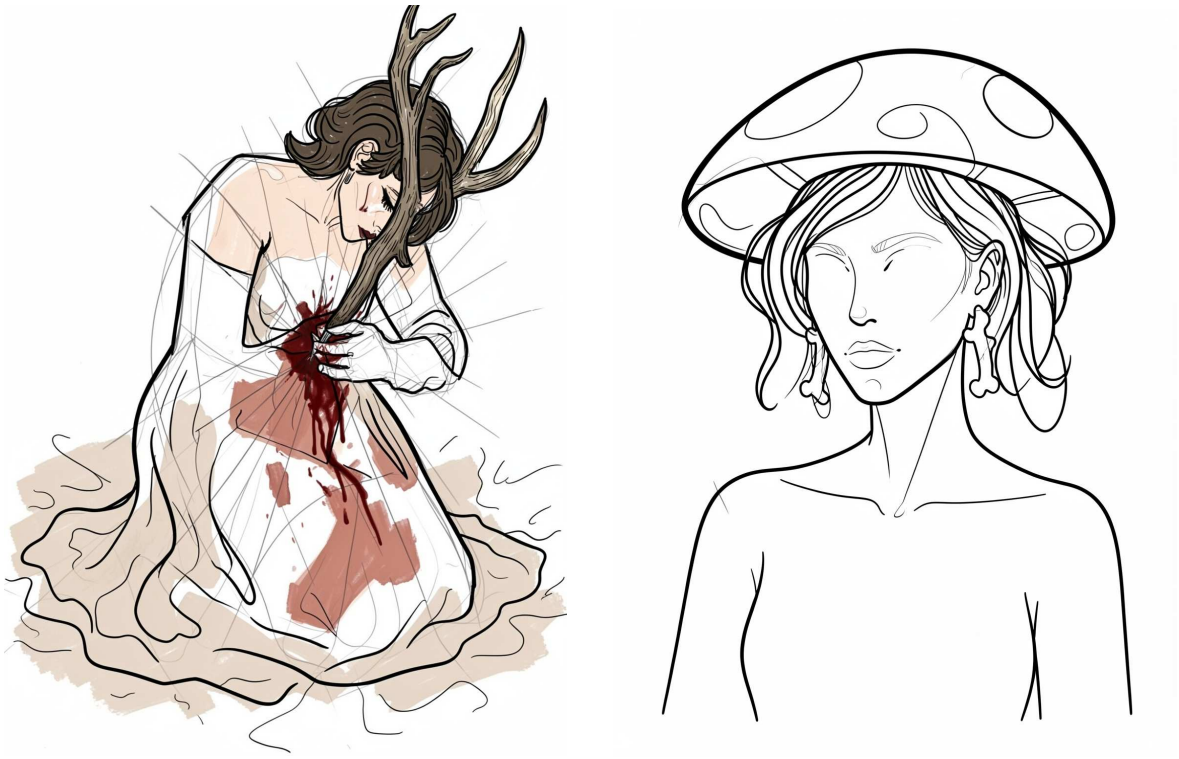
Unexpected mask image

Figure 7.2: Generated outputs that P1 treated as reference or inspiration material.

P1 also explained why finished references could still fit their process. In ordinary practice, they would place a finished reference image in Photoshop, lower its opacity, sketch over it, and add their own details. For P1, a finished-looking reference did not automatically remove their own creative input, because they treated references as material to adapt. This does not show that finished references carry no fixation risk, since the session did not evaluate delayed drawing from those images. It does, however, explain why the work-in-progress generation option did not fit P1's use case. The option had been added as a way to avoid overly finished outputs and leave more detail for the artist, but P1's response showed a different reference practice. They wanted the generated reference to have reliable anatomy and concrete visual information, because they would add their own rendering and details afterward. When the work-in-progress outputs were anatomically weak, they became less useful rather than more process-preserving:

> *"I need the anatomy to be right and that's what I use references for, so [...] that was very useless for me."*

The problem was not only visual style. P1 said that if an image was only a work in progress, style was less important because style is often the final rendering stage. The more serious issue was that the rough images did not provide the kind of anatomical reference they needed.



WIP self-stabbing pose

WIP mushroom woman

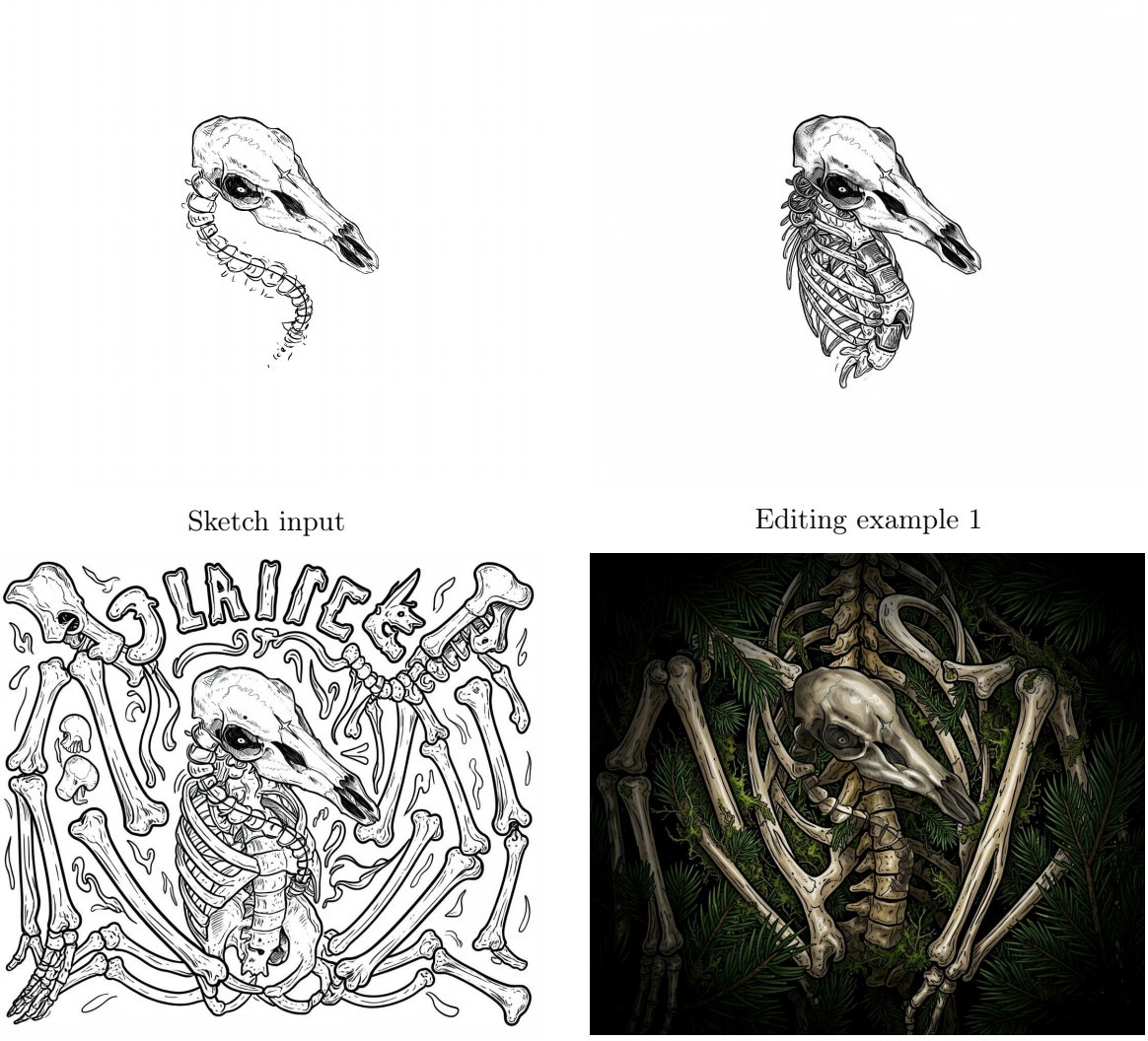
Figure 7.3: Work-in-progress outputs that P1 found unsuitable as direct references.

The earlier sketch-constrained ideation case remained unresolved. P1 initially tried the same unfinished sketch from the first session only briefly, because the first attempts did not appear useful. Near the end of the session, the researcher asked P1 to return to this case more explicitly to test whether the editing path had any potential for it. The task was difficult because P1 did not want only a local edit to a known object. They wanted the system to preserve the existing unfinished sketch while deciding what kind of content could be added around it. P1 framed this as an ideation problem rather than only an image-editing problem:

> *"[...] if you don't know what you want to draw at all, then it won't really solve the problem."*

Across several further prompts, the model did not produce a usable continuation. Some attempts changed the sketch substantially; others preserved parts of the original shape but did not use it as a meaningful structure, adding visually or semantically nonsensical material around it. Turning prompt mediation off for one attempt did not resolve the issue. One result was closer than the earlier attempts, but P1 still treated it as evidence that the model was struggling with the unfinished image rather than as a useful solution. They suggested that the sketch might not be a good test case because it was too ambiguous for the model to understand, and that a more recognizable unfinished figure might behave differently.

This finding should therefore be read narrowly. The attempt used only one unfinished sketch image, and P1 suggested that the model may have struggled because it was difficult to understand what the image showed. The session did not test the same function on other unfinished sketches because P1 did not have other comparable images available. In this case, the workflow did not become usable as a way to preserve an ambiguous sketch while inventing a meaningful continuation.



Sketch input

Editing example 1

Editing example 2

Editing example 3

Figure 7.4: Unresolved sketch-constrained editing attempt examples.

7.1.3 Personalization and Stylistic Fidelity

Personalization remained one of the strongest parts of the workflow for P1. They gave the personalization section of the questionnaire the maximum rating on every item. In the interview, they also compared the style recognition with the first session:

"Yeah, it's just as recognizable, but I think I like these more, but it might

be because of the prompting, so. It's hard to say like the style is just super recognizable."

P1 liked the outputs of the new workflow more, but did not know whether that came from the updated model, the LoRA, the prompt mediation, or the prompts used during the session. The evidence therefore supports the claim that personalization remained strong for P1, not that the new personalization pipeline alone caused the improvement.

The session also showed a narrower limitation of the redesigned generation path. As in the first workflow, a prompt could produce a more generic realistic image when the prompt did not clearly specify an illustrated or line-art style. In the first session, adding reference-image conditioning sometimes helped recover a recognizable style, although it usually also changed the composition. In the redesigned workflow, the researcher pointed out that editing mode could be used to try to change the style of such an output, and P1 attempted this once. The result was not satisfactory to them. This should not be read as showing that editing could never recover style, because the episode was not pursued further; after this point, P1 mostly avoided the issue by using style-relevant prompt wording.

P1 also recognized more specific traces of their work. In one detailed colored output, they said they could see similarity to one of the few colored images in the training set, especially in texture and highlights. They described the style reproduction as consistent, without one feature becoming too overpowering or too weak.

Image from dataset

Colored style detail

Figure 7.5: Dataset image and generated output with a similar recognized detail style.

The practical value of personalization was tied to reference use. Across the two sessions, P1 treated outputs outside their own visual style as less useful for reference work. In this session, this was especially clear for generic realistic outputs. Personalized images mattered because they fit P1's reference practice more closely than those outputs did. They made outputs more relevant and acceptable as source material.

7.1.4 Control, Privacy, and Ownership

P1's control, privacy, and trust ratings were again uniformly positive. The interview added little new evidence about privacy, but it confirmed that P1's position had remained stable. When the researcher noted that the same privacy questions had already been discussed, P1 said they did not think their opinion had changed, and when asked about cloud-based image tools, P1 said they would not use them for this kind of work.

Editing gave P1 more practical control over promising generated images because it allowed them to keep working from a specific output. However, this did not mean precise control over all image changes. The limits of prompt-based editing still shaped how reliably P1 could preserve or change an image.

Ownership was also discussed in a way that clarified the role of style. P1 first said that they felt they owned the generated images, but framed this partly as a preferred position rather than as a settled principle:

> *"Well, I feel like I own them [...] it's because I want to own them, but if they were trained on someone else's work then I wouldn't want to own them."*

P1 therefore still leaned toward owning the outputs, but the claim depended on whose work had shaped the model. When asked about outputs without the personalized style, they said:

> *"I would say I still own them, but they're just kind of useless [...] I don't feel so strongly about them if there's not the style transfer because [...] I feel like [...] my intellectual property is the style."*

This is close to the ownership position from the first session, but it is more explicit. For P1, personalization affected not only visual preference, but also whether an output felt connected to their own intellectual and creative material.

7.1.5 Summary of P1 Re-Evaluation

P1's second session suggests that the redesigned workflow addressed some of the main limits of the first workflow, but only in bounded ways. The clearest change was that prompting no longer appeared to be P1's main barrier. The local prompt mediator let

them begin from shorter, less model-specific instructions and seemed to make useful unexpected outputs more available. This made prompting less obstructive for P1, but it did not fully remove the difficulty of getting highly specific visual outcomes.

Editing added a second important path through the workflow. P1 could start from a generated image that already had useful qualities and try to steer it, rather than relying only on repeated regeneration from text. At the same time, editing was not the same as reliable structural control. It helped P1 work from a promising image, but it did not solve the more open-ended sketch-constrained ideation case. When P1 did not know what to add to an unfinished sketch, the session did not show that the redesigned workflow could preserve the ambiguous sketch and produce a meaningful continuation. It remains unclear whether different manual prompting, mediation instructions, or a less ambiguous sketch would have changed the result.

The work-in-progress option also showed a mismatch between a design response and P1's reference practice. It had been included to leave more room for the artist's own development, but P1 needed reference images with reliable anatomy and concrete visual information. Rougher outputs with broken structure were therefore less useful rather than more process-preserving for this participant. This finding should not be generalized beyond P1. Less finished outputs may still reduce fixation or support other artists, but in this session they did not fit how P1 uses references, and the same limitation may apply to artists who rely on references for concrete anatomical or structural information.

Overall, P1's response suggests that the redesign made the workflow more usable for the kinds of supported reference work that already fit their practice. Prompt mediation lowered the burden of writing model-specific prompts, editing made it possible to keep working from a promising generated image, and personalization kept the outputs close enough to P1's visual language to remain useful. These improvements did not make the prototype a fully controllable creative tool. The session still depended on researcher support, and the ComfyUI interface introduced practical frictions through small text, image history behavior, and technical node labels. The original sketch-constrained ideation case also remained unresolved.

Taken together, the session suggests that prompt mediation and editing made the redesigned workflow more useful for parts of P1's reference work, while sketch-constrained ideation and interface friction were the main areas for further improvement.

7.2 Results: Participant 3

P3's questionnaire responses after the redesigned workflow session were also very positive. Their System Usability Scale score was 100 out of 100. They gave the maximum rating to all creative utility items, and gave very high ratings for control, privacy, and trust.

The personalization responses were positive but slightly more nuanced. These scores are used descriptively. The session material is needed to interpret what the ratings meant in practice, especially because P3 had substantial prior experience with node-based creative software.

Metric	Score	Descriptive interpretation
Usability (SUS)	100 / 100	Maximum perceived usability in the questionnaire, interpreted together with P3's strong familiarity with node-based creative software.
Creative utility	5.0 / 5.0	Very high perceived usefulness, especially as inspiration, reference, source material, and possible support for downstream visual work.
Personalization	4.25 / 5.0	Strong preference for personalized outputs and clear recognition of style/material, with a more neutral response on authenticity to brand or identity.
Control, privacy, and trust	4.75 / 5.0	Very positive ratings for control, local/private use, and confidential work, with ownership rated slightly lower and discussed as collaborative.

Table 7.2: Descriptive questionnaire summary for P3's redesigned workflow session.

7.2.1 Usability, Prompt Mediation, and Workflow Control

P3 encountered the redesigned workflow for the first time, but their experience with the node-based interface of TouchDesigner made the ComfyUI interface easier to approach. P3 described the workflow as easy to use after a short initial orientation, saying that the initial learning period lasted only the first few minutes and that, after learning what to do, they were "just fine". They also connected this ease directly to TouchDesigner. Nodes could be turned on and off, or bypassed, in a way that was close to tools they already used. This made the prototype less intimidating for P3 and encouraged exploration. They said that, because of their experience with similar programs, they wanted to try things rather than being afraid to touch the controls.

This is important for interpreting the usability result. P3's response does not show that using the workflow would be equally easy for all artists. It shows that, for a participant already comfortable with node-based creative software, the redesigned workflow's visible controls and bypassable components could become understandable quite quickly. The session was still researcher-supported, but the debrief suggests that the workflow felt easy to operate once the main controls had been introduced. P3 also valued that the prototype remained stable during use. Near the end of the session, they

described the workflow as well optimized and contrasted this with their own experience of systems that fail or close during use.

Prompt mediation was one of the clearest supports for P3. They had used AI tools before, especially for text and local AI experiments, but reported very little use of visual generation. Their explanation was direct:

> *"It is because I cannot write the prompt. And I cannot describe myself or the things that I want to do in the way it will create something like this."*

The local prompt mediator addressed this barrier by translating shorter or less model-specific input into a more detailed image prompt. P3 inspected the mediated prompt and described the LLM support as something that had been missing from their previous experience with cloud image tools:

> *"This is maybe what I was missing, why I cannot use it so good or properly or efficient, so maybe this small thing just change[s] everything: my look on it [...] how I could use it."*

This interpretation should be cautious because several factors were present at once: personalization, the unfamiliar image model, high generation speed, and researcher explanation all shaped the session. Still, P3's comments support the claim that prompt mediation made visual generation more accessible for them. P3 also said that when they wanted something specific, the system held to the core prompt rather than moving to something completely different.

Speed also mattered for usability and creative momentum. P3 compared the workflow with slow Blender rendering, where rendering and checking an animation could take long enough for interest to fade. They said that when a process takes too long, they lose interest and their imagination is "done". In contrast, faster generation could produce several possible directions within one session:

> *"When I create all these in one hour [...] from here there are like five projects that I can start."*

For P3, speed was therefore not only a convenience feature. It affected whether the tool could support early exploration before motivation and attention moved elsewhere.

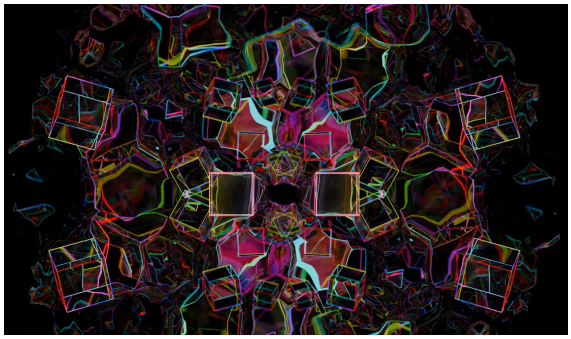
7.2.2 Creative Utility Across Generation and Editing

P3 did not treat the workflow mainly as a way to make final images. They treated generated and edited outputs as material that could be used in later work. This differed from P1's mainly reference-oriented use. P3 imagined placing images into TouchDesigner, extracting lines, removing color, using them as references or source files,

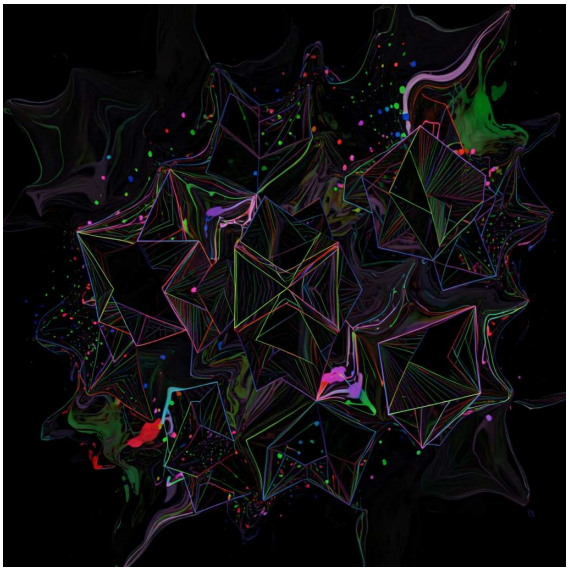
and post-processing them into further visuals. In the debrief they summarized this material use broadly:

> *"All the images that it created, I can use them all to create something new [...] some of the images can be used as an inspiration for me or it can be used as a specific type of art."*

This account explains why P3 could treat the session as creatively useful even without producing a finished artwork during the evaluation.



Example from dataset



Generated material example

Figure 7.6: Example from dataset and generated material.

Editing was useful in a similar way. P3 did not need every edit to preserve exactly what they had intended. When an edit changed the image in an unexpected way, they often treated the result as something that could still be continued from or processed elsewhere. This matched how they described working in TouchDesigner, where they could keep a core system in place while changing smaller parts around it. In one episode, an output did not do exactly what P3 had wanted, but they did not treat that as a failure:

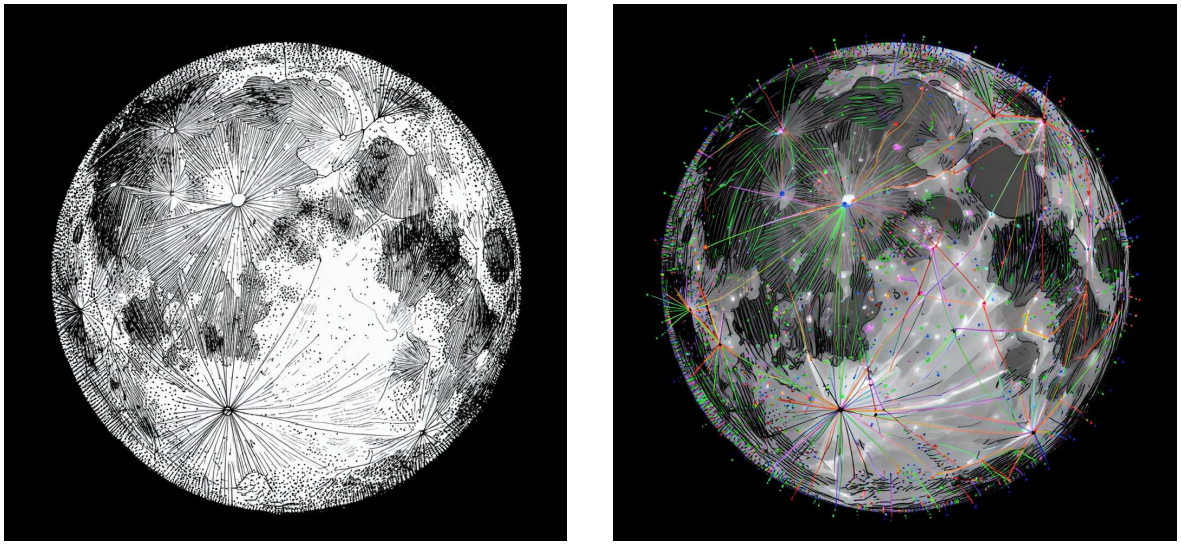
> *"I don't look at it like it's a bad thing [...] if I want to use this one, I can use and continue."*

This tolerance for drift shaped how editing was evaluated. For P3, control did not depend on the edit matching the instruction exactly; it depended on whether the changed image could still be used as material.

P3 also imagined editing at different stages of a project. They said they might use the workflow at the beginning of a process, but also near the end, for example to add

details to a finished 3D render or another completed image. This suggests that editing supported not only correction of generated images, but also possible augmentation of existing visual work.

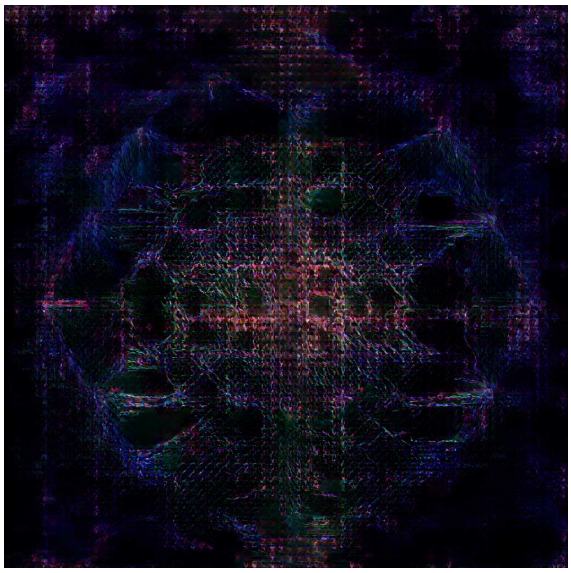
In a brief part of the session, P3 used a moon image to try changing the style through editing. The attempts produced mixed results. A sketch-like style applied more consistently across the moon, while the more personalized abstract dot-and-line style was harder to apply to the whole image. It kept the moon present and added visual elements around or over it rather than fully restyling the input. This suggests that editing could provide material for augmentation and restyling. However, the episode involved only a few attempts, and P3 moved on to another task rather than revisiting it. The session therefore does not show whether more sustained prompting or iteration could have produced more reliable style transfer while preserving the input structure.



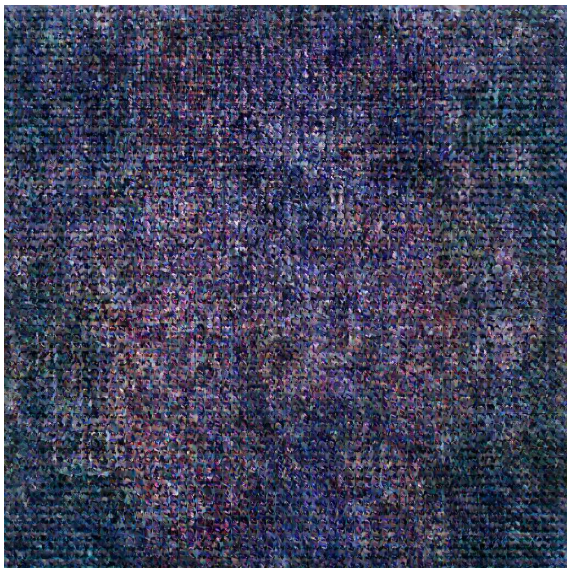
Sketch-style edit Personalized style attempt

Figure 7.7: Examples from P3's brief moon style-editing episode.

P3 also explored LoRA strength as a creative control. The default setting was intended as a reasonable personalization strength, but the interface allowed the participant to change it freely. P3 increased the strength beyond the usual range and found some of the resulting artifacts useful. This is a P3-specific finding because their practice made patterns, textures, and visual noise potentially valuable. In this case, exposing control over personalization strength allowed P3 to discover material uses that a more constrained interface might have hidden.



High-strength LoRA artifact



High-strength LoRA artifact

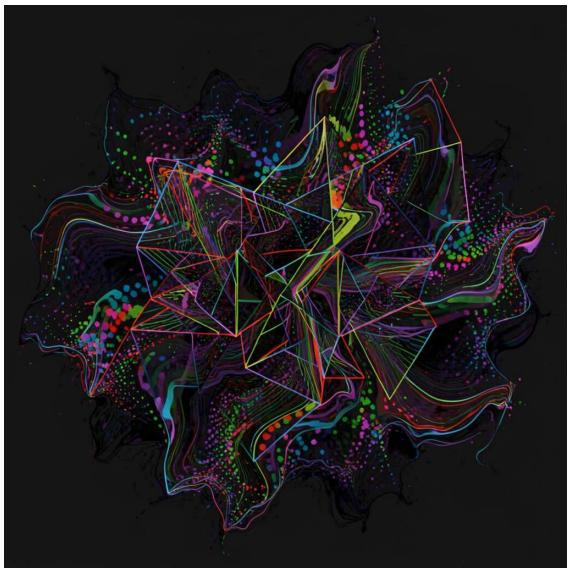
Figure 7.8: Examples of high LoRA-strength outputs that P3 treated as potentially useful material.

7.2.3 Personalization and Stylistic Fidelity

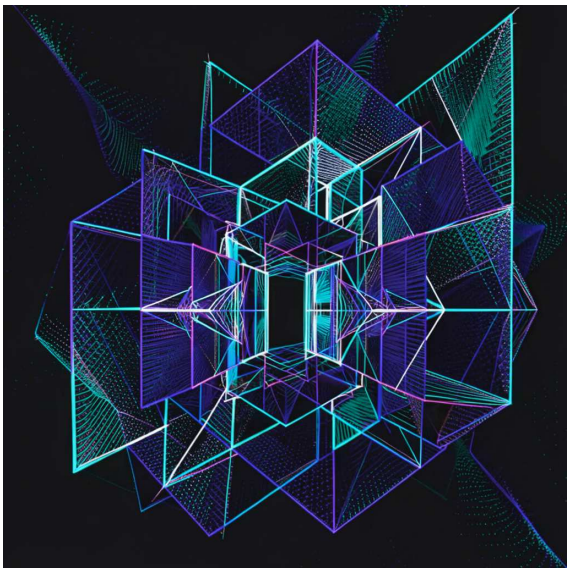
Personalization was one of the main reasons the workflow felt useful to P3. During the session, they repeatedly recognized features from their own material in the generated images, including line patterns, sketch-like marks, geometric forms, and fluid or abstract visual traces. In the debrief, they said the system captured the trained style "a lot". When explaining why this mattered, they linked the training material to visual preference and to having outputs they could continue exploring:

> *"Because it was trained on my style or the things that I do. It was good because I can see like the style I like."*

P3 made a similar point when comparing the workflow with cloud tools. They described the personalized outputs as more usable because they liked them more than images from cloud services they had used before. In this sense, recognizing their own material mattered not only as evidence of style transfer, but also because it made the outputs more appealing to work with.



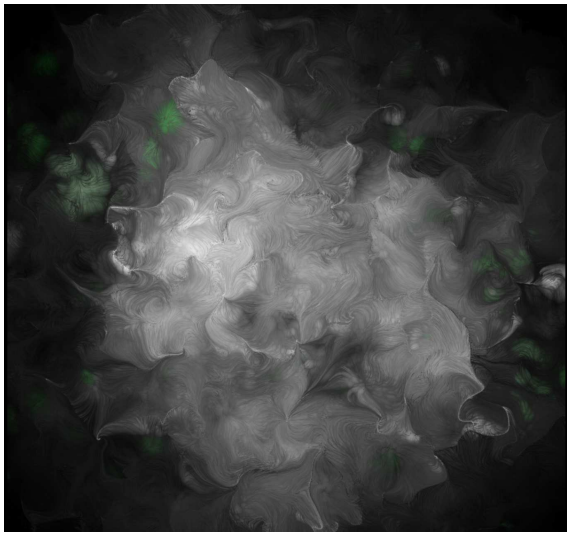
Personalized output example



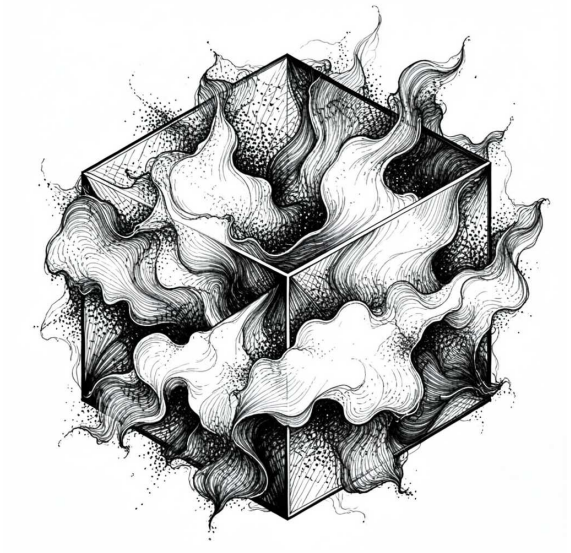
Personalized output example

Figure 7.9: Examples of personalized outputs recognized by P3.

P3 also explored prompts that described a sketch-like style while the personalized model still introduced elements associated with their own dataset. In this episode, the output appeared as a synthesis between the prompted style direction and the learned visual material.



Dataset example



Generated style synthesis example

Figure 7.10: Example of a dataset image and a generated output combining P3's personalized visual material with a sketch-like prompt direction.

P3 also noted in the interview that repeated style could become limiting if they wanted something completely new. Their response was practical rather than negative. They expected they could write the prompt differently or change the look afterward.

For P3, personalization was not a complete capture of artistic identity. It was a useful way to make outputs feel closer to their own material and more usable as starting points.

7.2.4 Control, Privacy, and Ownership

P3's acceptance of AI in their own workflow was high. They did not frame the workflow as a replacement for artists. Instead, they treated AI as a normal tool and described outputs as inspiration or material that would still be remade, combined, or processed through their own work. Their position was summarized in one short comment:

> *"It will not replace me. It will not just magically do stuff. It will just give me inspiration."*

This tool framing helps explain why imprecise or artifact-heavy outputs could still be valuable. The workflow was not expected to complete P3's work by itself.

P3's privacy response was also positive, but it had a specific meaning. They strongly rated private operation as increasing their willingness to use the system, and they strongly agreed that they would be comfortable using it for confidential or unreleased professional work. In the interview, they emphasized personal expression rather than only professional secrecy. Private operation mattered especially when the work involved personal experiences, emotions, or parts of their life:

> *"When I want to create work where I put part of myself [...] where I just explaining or just recreating for example my emotions, what I feel [...] it would be better if it stays in there."*

P3's response to local operation was also practical. Early in the session they asked about the technical requirements for running the workflow, noted that they had access to an RTX 3090, and imagined running the system on a home PC while accessing it from a laptop. For P3, local use was therefore not only a privacy preference. It was also something they could connect to their existing hardware and prior experience with local AI systems.

Ownership was more complex. P3 did not treat AI-assisted output as purely theirs or someone else's. Instead, they compared it to film production, where different people contribute to a finished work:

> *"It's hard to say because it's work of the all of them. So I think it's like a teamwork, maybe."*

If they used the system in work they published or presented, P3 said they would acknowledge that the AI or workflow had been used and credit the creator of the model

or system as part of the work. At the same time, they felt more ownership than they would with generic cloud tools because the model was trained on their own material. This makes P3's ownership position distinct from a simple claim that the images were solely theirs. Personalization increased their sense of connection to the outputs, while ownership and credit were still understood as shared.

P3 also compared the session positively with cloud-based AI tools. They described the experience as much better, easier, simpler, and faster than their previous cloud-tool use. The comparison likely reflected the combination of local prompt mediation, personalized outputs, faster iteration, visible controls, and the supported session context.

7.2.5 Summary of P3 Results

P3's session adds a redesigned-workflow case centered on visual material for further digital work. P3 approached the workflow from a practice that included node-based tools, post-processing, and local AI experimentation. Fast generation and prompt mediation made the workflow easier to explore, while personalization and editing gave P3 material they wanted to keep working from. The visible controls, possible self-hosting, and similarity to P3's existing node-based tools also made the prototype feel closer to systems they already understood.

P3 valued the outputs as material for later work, including inspiration, references, patterns, source images, post-processing inputs, and starting points for further TouchDesigner and Blender work. This also changed how editing was evaluated. For P3, an edit did not need to be exact to be useful. Unexpected changes and artifacts could become material to continue from.

P3's privacy and ownership responses were also shaped by their own practice. Private operation mattered especially for personal work, while ownership was understood as collaborative rather than purely individual.

P3's ease of use was shaped by their prior experience with node-based tools, and the prototype still depended on researcher support. Longer independent work was not tested. Editing supported restyling and augmentation, but the brief moon episode showed mixed success when the goal was to apply a more participant-like abstract style while preserving the input image. In this session, the workflow appeared useful for P3's exploratory material-based practice, while precise spatial or structural control remained uncertain.

7.3 Cross-Case Synthesis

The two redesigned-workflow sessions involved different artistic practices, but they did not point to completely separate kinds of usefulness. P1 mainly wanted references

and help with art block, while P3 looked for images that could become inspiration, source material, or inputs for later post-processing. In both cases, the workflow was not treated mainly as a system for producing finished images. What mattered was that it produced something the participant could react to, transform, draw from, or carry into another process. The workflow therefore gave both participants ways to get usable material out of a short supported session, while still leaving limits around interface polish, structural control, and longer-term independent use.

7.3.1 Usability and Workflow Control

Both participants gave very positive usability ratings. P1's SUS score was 90 out of 100, and P3's was 100 out of 100. These scores fit the session material. Both participants were able to use the redesigned workflow during the evaluation, explore its main functions, and describe the experience as easy overall.

Across both sessions, the workflow was straightforward to use, with minor frictions rather than major usability barriers. P1 had already used the earlier ComfyUI workflow, while P3 could relate some parts of the graph to TouchDesigner, especially bypassing nodes and trying different settings. Researcher explanation was also part of both sessions. Within that context, both participants could use the redesigned workflow to explore generation, prompt mediation, editing, and personalization.

The local prompt mediator also made the workflow easier to start using. For P1, it made prompting less effortful both for quick simple prompts and for cases where a more model-effective formulation was needed, even though outputs still sometimes required iteration. For P3, mediation addressed a prior difficulty with visual AI tools, where they found it difficult to describe an intended image in a form that the system could use. In both cases, the mediator helped participants get started without making prompt writing the whole task.

The remaining usability issues were mostly about navigation, visibility, and terminology. P1 found the image-history behavior annoying when trying to continue editing the same image, and also noted small text and technical labels such as LoRA. Prompt mediation also did not remove the need for visual judgment, iteration, or some awareness of how style wording affected the output. They remained secondary to the overall ease of use reported in the sessions. The redesigned workflow was easy enough to use for supported creative exploration, while longer unattended use remained outside the evaluation.

7.3.2 Generation and Editing as Material Support

In both sessions, outputs were treated as material for later work. P1 wanted references and art-block support, and during the session editing made some generated images more

useful as possible references. P3 wanted inspiration, references, source images, patterns, and material for further work in tools such as TouchDesigner or Blender. These are different creative workflows, but they share an important structure. In both sessions, the value of the workflow came from producing material that could be continued from rather than from replacing the participant's own creative process.

The kind of material needed was different. P1 needed concrete visual information for reference use, especially usable anatomy and a recognizable visual style. This is why the work-in-progress option did not fit P1's reference practice when it produced rougher images with weaker anatomy. P3 did not take up the work-in-progress option after it was introduced, so the session does not show whether that mode would have been useful for their practice. P3 instead talked about how looser outputs, including patterns, artifacts, and unexpected visual changes, could be processed, layered, or used as starting points in another digital system.

Editing showed practical value in both cases. For P1, editing made it possible to keep working from a generated image that already had promising qualities. In the pose-editing episode, P1 got quite far toward a substantial transformation of the generated image, even though the final result was not exactly their full intention. This gave P1 another option besides prompt-only regeneration, where useful parts of an image could be lost when the prompt was changed. For P3, editing was useful as a way to continue from an image, augment it, or imagine adding details to existing visual work. Their ideation use differed from P1's reference and sketch-oriented use, but both participants found ways to use editing as a continuation path.

Editing was less clear in two places. P1's earlier sketch-continuation case remained unresolved in the redesigned session: for the one ambiguous sketch image that was tested, the workflow did not produce a useful continuation. P3's brief moon-editing episode showed mixed success when trying to apply a more participant-like abstract style while keeping the moon as the input structure. In short use, both participants already got meaningful material from the workflow, while longer real-project use would be needed to understand how consistently those useful moments hold across tasks.

7.3.3 Personalization and Relevance

Personalization remained important in both cases. P1 again recognized their style strongly, and personalized outputs were more useful as references because they were closer to P1's own visual language. For P1, this mattered because the images were being judged as references. Images that felt disconnected from their way of drawing were less useful for that purpose.

P3 also recognized their own material in the outputs and preferred the personalized results over more generic outputs. For P3, personalization made the images more

appealing and more usable as inspiration or source material. They noted that repeated style could become limiting if they wanted to make something completely new, and expected that they could change the prompt or alter the look afterward. In the workflow itself, the style path could also be bypassed when it was not wanted.

7.3.4 Privacy, Ownership, and Acceptance

Both participants responded positively to the private and local character of the workflow, but the value of privacy differed. For P1, privacy remained mainly protective. Their concern was that portfolio material and stylistic information should not be exposed to a cloud service or absorbed into a larger external system. The redesigned session did not change this position; it confirmed that private local use was part of why the workflow was acceptable for this kind of work.

For P3, privacy was connected more directly to personal expression and practical local use. They described private operation as valuable when creating work that involved personal experiences, emotions, or parts of their life. They also discussed local use in relation to available hardware and prior experience with local AI systems. This made local operation seem plausible as part of their own setup.

Participants also interpreted ownership differently across the two cases. P1 leaned toward feeling that the user owns outputs when the system is trained on their own work, and their sense of ownership was stronger when the output carried their style. P3 framed ownership more collaboratively. They felt more connection to the outputs than they would with generic cloud tools because their own material was involved, but they also described the model, workflow, and tool builder as part of the production process. In both cases, using the participant's own material mattered for how they thought about ownership and credit, but it did not lead to the same position.

Taken together, the redesigned-workflow sessions suggest that the redesign addressed several issues raised by the first evaluation. Prompt mediation reduced the burden of writing model-effective prompts, editing added a way to continue from and transform promising images, and personalization remained central to whether outputs felt relevant. The two cases also show that this value was not limited to one exact practice. P1 and P3 worked differently, but both used the workflow as a way to produce material for further creative work. The remaining boundaries are also clear: the workflow was still used in short supported sessions, the ComfyUI interface retained practical frictions, longer project use was not tested, and the sketch-continuation case remained unresolved. Broader implications of these findings are discussed in the following chapter.

Chapter 8

Discussion

Across the two design cycles, the workflow's value was clearest when generated images became part of an ongoing creative process. The participants did not treat the system mainly as a replacement for making images themselves. They used it to produce material they could react to, explore, or adapt.

The findings should be read in relation to the study's exploratory design. The sessions were short and researcher-supported. Within those limits, the evaluations show how particular design choices mattered in use. They also suggest that image quality alone may be too narrow a way to evaluate the creative utility of a visual generative AI system.

8.1 Generated Images as Creative Material

Across the sessions, generated images were useful when participants could place them within an existing or imagined workflow. They worked as references, inspiration, style experiments, prompts for further thinking, or source material for another step. Their creative value was therefore not reducible to how polished the image looked. This matches the utility gap described in the literature, where artist-facing image generation is often valuable because it makes possible directions visible within an ongoing process (Ko et al., 2023; Johnston and Thue, 2024). Good images were still valued, but the important question was whether an output opened up a useful next move in the participant's creative process. An image could still have creative value when it was imperfect or not exactly what was intended, as long as it contained something the participant could use or respond to.

The work-in-progress generation option was based on the same concern with provisional material. It was added because finished personalized examples can be ambivalent, and because fixation research suggests that polished examples can narrow later exploration (Wadinambiarachchi et al., 2024). The evaluation of the redesigned workflow

did not test this mode enough to say much about its value. P1 tried it briefly, and in that context the rougher outputs did not fit their reference use because they lacked the anatomical information they needed. The mode may still support other kinds of ideation, but this remains mostly open from these sessions.

Generation speed also mattered because the workflow was used through repeated trials rather than through single outputs. P3 made this most explicit by connecting fast generation with sustained interest and with finding several possible project directions in a short session. The other sessions were less direct about speed, but they also depended on repeated trial, selection, and adjustment. In a workflow used for exploration, faster generation makes that repeated sampling more practical and can increase how much usable material is available within the same session.

8.2 Control Across Generation and Editing

The first evaluation showed that a workflow can be usable in a session while still leaving participants uncertain about control. They could operate the prepared interface, but they could not always tell how prompts, references, personalization settings, and conditioning paths would affect the result. The problem was not only how to produce an image. It was how to keep the important part of an intention stable while changing something else.

In the first evaluation, participants sometimes worked around prompt difficulty by using an external LLM, with mixed results. The redesign brought this support into the local workflow. This did not remove the need for judgment or iteration, and it cannot be separated completely from the other changes in the redesigned system. Still, the sessions suggest that mediation made generation easier to start from shorter or less model-specific prompts. Its value was also tied to visibility. The mediated prompt remained inspectable and bypassable, so participants could see how their input had been rewritten and decide whether to use it.

The other control problem appeared after a useful image already existed. Changing the prompt could lose the composition, pose, texture, or mood that made the earlier output worth keeping. Editing gave participants a way to keep working from that image instead. The edits were not always successful, but the interaction still allowed another attempt from the same starting point. That retryable continuation was important because it shifted the workflow away from prompt-only regeneration.

The limits of this redesign appeared when participants needed more precise structure. Prompt-based editing still works through language. It does not provide the same kind of constraint as a mask, pose guide, or other spatial control. The evidence for this limit is narrow. The sketch-continuation attempt used one ambiguous image, and the restyling

episode was brief. These cases therefore do not show how the workflow would perform across structure-preserving tasks more generally. They do show that the redesigned workflow did not yet provide clear evidence of reliable support for this kind of control.

8.3 Personalization, Style, and Ownership

All participants consistently valued the personalized outputs. The recognizable style could make images easier to relate to, easier to consider as reference, or more appealing as source material. It could also make the use of generated images feel more connected to the artist's own contribution. In this sense, style fidelity was one of the clearest sources of value in the workflow.

The evaluations also leave open how much style control an artist-facing system should expose. The first workflow's richer personalization controls were harder to understand, but they also exposed different ways of applying participant-specific visual material. Reference-based conditioning could affect composition and style differently from the LoRA path, and could apply stylistic material in situations where the LoRA alone was weaker. It also opened up possibilities for style synthesis when different personalization paths brought different visual styles into the same generation. The redesigned workflow appeared easier to work with in the supported sessions, but it also narrowed these forms of control. The high-strength LoRA episode in the redesigned session shows the same point in a smaller way. Because the personalization control was exposed, P3 could discover an unexpected source of inspiration in artifacts that a more restricted interface might have prevented.

Personalization also shaped feelings of ownership. Outputs trained on and visibly connected to a participant's own material could strengthen a sense of ownership, especially when style was treated as part of the participant's intellectual contribution. This was clearest for P1. P2 recognized the personalized style but did not care about ownership of the generated images. P3 described the process more collaboratively. The small set of responses fits the broader uncertainty around ownership, style, training data, and human contribution (Lovato et al., 2024; Khan et al., 2025; Messer, 2024).

8.4 The Value of Local Deployment

Local deployment mattered because it affected what kinds of artistic material participants felt able to bring into the workflow. Material tied to style, professional work, or personal expression carried different stakes than ordinary test inputs. Keeping the workflow inside the researcher-controlled environment made some forms of exploration feel more acceptable, because prompts, input images, personalized model files, and

outputs did not have to pass through an external service.

This connects the evaluation to concerns in the literature around consent, dataset reuse, style mimicry, and provider control (Lovato et al., 2024; Jiang et al., 2023; Shan et al., 2023; Kabir et al., 2024). In this context, local deployment changed where trust had to be placed. A cloud service would require artists to rely on a provider's terms, retention policies, and future data practices. In the evaluated workflow, control over data handling instead sat with the administrator of the local system, which in this case was the researcher. Participants were told how their material would be stored, processed, and kept within the local setup, and agreed to take part on that basis. This still required trust in the researcher, but the data path was explicit and bounded by the local setup. If the artist also administered the system, the same arrangement could give them direct control over the data throughout the workflow.

The local workflow was also practical in a hardware sense. The redesigned workflow ran on the RTX 4080 Super setup used for the project, and the same setup was used to train the LoRA modules. This shows that consumer-grade hardware could support both local generation and LoRA training for the workflow. At least one participant already had access to comparable hardware in their own creative context, so running a similar workflow locally was plausible from the hardware side for that case. Independent deployment would still require more than owning suitable hardware. Participants did not install software, train LoRAs, or maintain the workflow over time.

Local control also depends on the practical work of setup and maintenance. In this thesis, the researcher handled that work. For an artist-facing system, local deployment would support agency only if artists could manage the workflow well enough to keep it usable and understand how their files and models were being handled.

8.5 Implications for Artist-Facing Workflow Design

The evaluations suggest several design implications for generative AI systems. One implication is that generation should not be treated as a self-contained step whose value is decided only when an image appears. In these sessions, the value of an output often depended on how it could be taken up afterward. It could support ideation, serve as a reference, or become source material for another part of a process. This makes it important for the workflow to support what happens after a promising image appears: keeping it available, comparing it with alternatives, adapting it, or moving it into another step. The custom ComfyUI support added for the redesigned workflow was built around this kind of continuity, making earlier outputs available for editing and linking saved generations back to their original and enhanced prompts. By placing editing next to generation, the redesigned workflow made it easier to continue from

an output that already had useful qualities. This mattered both when an image was promising but needed adjustment and when an already useful image invited variation or further exploration. Editing gave the participant another path from that image, instead of requiring a return to prompt-only regeneration. The sessions were less conclusive for cases where an artist needs to preserve the structure of an existing image. The brief attempts in this direction were too limited to show whether prompt-based editing could support those uses reliably.

Prompt mediation is one of the clearer design implications of the thesis. Across the evaluations, prompting appeared as a barrier between participants' own ideas and prompts that produced useful images. The usefulness of mediation was visible across the sessions, both in participants' responses to the integrated mediator and in the earlier use of external LLM support. This suggests that systems for artists may benefit from prompt support that stays close to the generation workflow rather than requiring a separate tool. Visibility still matters within this design. Participants occasionally inspected the rewritten prompt, and the mediator was available as something they could use, ignore, or learn from. Keeping it inspectable and bypassable made it closer to an additional tool in the workflow than to a hidden correction applied in the background.

Personalization was one of the clearest sources of relevance in the workflow. Across the sessions, participants valued outputs that carried a recognizable connection to their own visual material. This made generated images easier to relate to and, in some cases, easier to imagine as reference, inspiration, or source material. The value was not only that the images resembled the participant's style, but that this resemblance made them easier to place within the participant's own practice. Personalization can therefore affect whether generated images feel relevant enough to be considered as material for creative work.

The evaluations also suggest that exposed controls can create creative value, even when they add complexity. The first workflow exposed more ways of applying and combining participant-specific visual material, while the redesigned workflow narrowed those options and appeared easier to work with in the supported sessions. The trade-off matters because controls that are hidden or reduced to a single preset may also remove opportunities for exploration. When participants could try different strengths, references, or personalization paths, unexpected effects sometimes became useful as inspiration or as material for further work. The implication is not that more complexity is always better, but that systems intended for creative practice need to decide carefully which controls remain available, especially for users with enough technical fluency to turn them into creative value.

Taken together, these findings point against treating image quality as a proxy for creative utility. Image quality matters, but it mainly describes the output itself. In these sessions, participants recognized that generated images could help develop an

idea, serve as references, or provide material for later transformation. These possible uses depended on more than visual quality. They also depended on whether the system provided enough control, made important choices visible enough to understand, supported relevant personalization, and let useful material be kept and worked with. A system can therefore look strong by output quality while still offering limited creative support.

The level of privacy a system provides also affects what artists may feel able to bring into it. This is especially important for workflows built around personal style, input images, and trained model files. In such cases, data handling shapes whether artists can use sensitive or personally meaningful material without treating it as something handed over to an external service. The local workflow explored one way of addressing this concern by keeping those inputs and model files under the control of the artist or administrator, rather than making provider trust a condition of use.

8.6 Limitations and Future Work

The exploratory design limits how far the findings can be taken. The participant sample was small and purposively selected, and it covered only a limited range of artistic practices, styles, and technical backgrounds. The sessions produced detailed data about a few cases, but not a broad account of how different artists would use a private, personalized generative workflow. The evaluations were also short and researcher-supported. Participants used a prepared workflow, but did not install it, maintain it, train LoRAs independently, or use it across a longer project. The high usability scores should therefore be read as evidence that participants could operate the prepared workflow during the sessions, not as evidence that the same workflow would be easy to adopt independently.

P1's case also needs to be read with their personal familiarity with the researcher and passive familiarity with the project in mind. This may have shaped both the design problems the project foregrounded and P1's comfort, trust, or willingness to criticize the workflow during the evaluations.

Future work could broaden the participant sample, reduce researcher support, and examine longer use in artists' own projects. It could also test the less resolved parts of the workflow more directly: work-in-progress generation and structure-preserving editing. It could also evaluate a version that artists try to set up and use more independently, where local deployment and LoRA training are made easier through clearer packaging, a guided installation process, and a simpler training workflow.

Conclusion

This thesis examined how a private, stylistically personalized generative AI workflow can support visual artists. The work was situated in human–AI creative practice, where image generation is not only a question of producing high-quality outputs, but also of whether those outputs can be used within an artist's own process. The aim was to design and evaluate a local, stylistically personalized workflow, and to examine its usability, creative utility, stylistic fidelity, and relation to control, trust, privacy, and acceptance. After the first evaluation, the workflow was redesigned to add prompt support and prompt-based editing. Within the limits of a small exploratory study, these aims were fulfilled: the initial workflow was implemented, the redesign responded to findings from the first evaluation, and both versions were examined in situated sessions with artists.

The central finding is that the workflow was most useful when it supported continuation rather than replacement. Participants did not treat the system primarily as a substitute for their own creative practice. They valued generated and edited images as references, inspiration, source material, and starting points for further work. The creative value of an output therefore depended less on whether it could stand alone as a finished image and more on whether it could support the participant's next move.

The first workflow showed the importance of local deployment and participant-specific personalization, but it also revealed the limits of generation alone. Participants could recognize traces of their own visual material in the outputs, which made the images feel less generic and easier to relate to. At the same time, they wanted to transform existing sketches or project material while preserving what mattered in them. Prompting also remained a practical barrier. These findings motivated the redesign toward a workflow that kept local personalized generation, but added prompt mediation and prompt-based editing in the same environment.

The redesigned workflow showed that this combination can support several forms of creative material use. For P1, it was useful for reference generation, brainstorming, and continuing from images that already had promising qualities. For P3, it was useful as a source of inspiration, patterns, visual material, and possible inputs for later digital processing. These cases differed, but both point to the same interpretation: the image mattered as something the participant could take up within an existing practice, not as

a finished substitute for that practice.

Prompt mediation addressed a practical translation problem. Participants often had visual intentions that were not easy to express in prompts the model would respond to, so the mediator reduced part of the burden of turning those intentions into model-effective instructions. It also made prompt exploration less laborious, because participants could start from a rough intention and use the mediator to develop more specific prompt wording.

Editing added a continuation path by letting participants work further from images that already had useful qualities, instead of returning only to prompt-based regeneration.

Personalization remained important throughout the evaluations. Personalization in the workflow made some outputs more recognizable and acceptable as material for further work, and it supported exploration through combinations of trained style, prompted style directions, reference material, and stronger personalization settings. This does not mean that the model captured artistic identity in a full human sense. It means that recognizable traces of a participant's material could make outputs easier to relate to and more useful as starting points. Personalization also affected ownership unevenly. For P1, outputs linked to their own style strengthened a sense of ownership, while for P2 and P3 ownership was more ambiguous or collaborative.

Local deployment was also part of the creative support rather than only a technical choice. Keeping prompts, images, LoRA files, and outputs inside a bounded local setup affected what participants felt able to bring into the workflow. This mattered especially for portfolio material, client work, and personally meaningful creative content. The privacy value was not identical for every participant, but across cases data handling shaped what kinds of material could be used comfortably.

The contribution of the thesis is not a general claim that local personalized AI systems are ready for broad artist adoption. The study was small, exploratory, and researcher-supported. Participants did not install the workflow, maintain the software, train LoRAs independently, or use the system over a long project. The contribution is a design and evaluation account showing how privacy, personalization, prompt mediation, and editing can work together to make generated material more relevant, usable, and acceptable for artists. The findings suggest that artist-facing generative AI systems should be evaluated less as sources of isolated images and more as parts of continuing creative processes, where utility depends on the relation between an output and what the artist can do next.

Bibliography

Abuzuraiq, A. M. and Pasquier, P. (2025). Explainability-in-Action: Enabling Expressive Manipulation and Tacit Understanding by Bending Diffusion Models in ComfyUI.

Baek, S., Im, H., Ryu, J., Park, J., and Lee, T. (2023). PromptCrafter: Crafting Text-to-Image Prompt through Mixed-Initiative Dialogue with LLM.

Black Forest Labs (2024). FLUX.1-dev. Hugging Face model card. `https://huggingface.co/black-forest-labs/FLUX.1-dev`.

Black Forest Labs (2026a). FLUX.2 Image Editing. Official documentation. `https://docs.bfl.ai/flux_2/flux2_image_editing`.

Black Forest Labs (2026b). FLUX.2-klein-9b-kv-fp8. Hugging Face model card. `https://huggingface.co/black-forest-labs/FLUX.2-klein-9b-kv-fp8`.

Black Forest Labs (2026c). FLUX.2 Overview. Official documentation. `https://docs.bfl.ai/flux_2/flux2_overview`.

Black Forest Labs (2026d). Prompting Fundamentals. Official documentation. `https://docs.bfl.ai/guides/prompting_guide_t2i_fundamentals`.

Black Forest Labs (2026e). Prompting Guide – FLUX.2 [klein]. Official documentation. `https://docs.bfl.ai/guides/prompting_guide_flux2_klein`.

Black Forest Labs, Batifol, S., Blattmann, A., Boesel, F., Consul, S., Diagne, C., Dockhorn, T., English, J., English, Z., Esser, P., Kulal, S., Lacey, K., Levi, Y., Li, C., Lorenz, D., Müller, J., Podell, D., Rombach, R., Saini, H., Sauer, A., and Smith, L. (2025). FLUX.1 Kontext: Flow Matching for In-Context Image Generation and Editing in Latent Space.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., and McClelland, I. L., editors, *Usability Evaluation in Industry*, pages 189–194. Taylor & Francis.

Cao, P., Zhou, F., Song, Q., and Yang, L. (2026). Controllable Generation With Text-to-Image Diffusion Models: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(4):4771–4791.

Chang, Z., Koulieris, G. A., Chang, H. J., and Shum, H. P. (2026). On the design fundamentals of diffusion models: A survey. *Pattern Recognition*, 169:111934.

Chompunuch, S. and Lubart, T. (2025). AI as a Helper: Leveraging Generative AI Tools Across Common Parts of the Creative Process. *Journal of Intelligence*, 13(5):57.

Comfy-Org (2026). ComfyUI. GitHub repository. `https://github.com/Comfy-Org/ComfyUI`.

Doshi, A. R. and Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28):eadn5290.

Fan, X., Wu, Z., and Wu, H. (2025). A Survey on Pre-Trained Diffusion Model Distillations.

Fang, S. (2024). A Comprehensive Survey of Text Encoders for Text-to-Image Diffusion Models. *EAI Endorsed Transactions on AI and Robotics*, 3.

Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A. H., Chechik, G., and Cohen-Or, D. (2022). An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion.

Gale, N. K., Heath, G., Cameron, E., Rashid, S., and Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, 13(1):117.

Garcia, M. B. (2025). The Paradox of Artificial Creativity: Challenges and Opportunities of Generative AI Artistry. *Creativity Research Journal*, 37(4):755–768.

ggml.org (2026). llama.cpp. GitHub repository. `https://github.com/ggml-org/llama.cpp`.

Ghori, I., Karim, K., and Alkawadri, D. (2025). GenAI-Driven Image Generation Pipeline for Sustainable Garment Design and Waste Reduction in Fashion Production. *Proceedings of the AAAI Symposium Series*, 6(1):218–226.

Google AI Edge Team (2026). Bring state-of-the-art agentic skills to the edge with Gemma 4. Official developer blog. `https://developers.googleblog.com/bring-state-of-the-art-agentic-skills-to-the-edge-with-gemma-4/`.

Ham, S.-J. and Park, C.-S. (2026). Efficient and Controllable Image Generation on the Edge: A Survey on Algorithmic and Architectural Optimization. *Electronics*, 15(4):828.

Hao, Y., Chi, Z., Dong, L., and Wei, F. (2023). Optimizing prompts for text-to-image generation.

He, Y., Liu, Z., Chen, J., Tian, Z., Liu, H., Chi, X., Liu, R., Yuan, R., Xing, Y., Wang, W., Dai, J., Zhang, Y., Xue, W., Liu, Q., Guo, Y., and Chen, Q. (2024). LLMs Meet Multimodal Generation and Editing: A Survey.

Hei, N., Guo, Q., Wang, Z., Wang, Y., Wang, H., and Zhang, W. (2024). A User-Friendly Framework for Generating Model-Preferred Prompts in Text-to-Image Synthesis.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising Diffusion Probabilistic Models.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models.

Huang, Y., Huang, J., Liu, Y., Yan, M., Lv, J., Liu, J., Xiong, W., Zhang, H., Cao, L., and Chen, S. (2025). Diffusion Model-Based Image Editing: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4409–4437.

Jiang, H. H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., Hanna, A., Flowers, J., and Gebru, T. (2023). AI Art and its Impact on Artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pages 363–374, Montréal QC Canada. ACM.

Johnston, H. and Thue, D. (2024). Understanding Visual Artists' Values and Attitudes towards Collaboration, Technology, and AI. In *Graphics Interface*, pages 1–9, Halifax NS Canada. ACM.

Kabir, A. I., Mahomud, L., Al Fahad, A., and Ahmed, R. (2024). Empowering Local Image Generation: Harnessing Stable Diffusion for Machine Learning and AI. *Informatica Economica*, 28(1/2024):25–38.

Kambur, H. and Dolunay, A. (2024). A research on copyright issues impacting artists emotional states in the framework of artificial intelligence. *Frontiers in Psychology*, 15:1409646.

Kaur, A., Singh, S., and Kaur, N. (2026). A Survey of Text-to-Image Generation: From GANs to Diffusion Models, Datasets, Evaluation Metrics and Key Challenges.

Khan, M. A., Mikalonytė, E. S., Mann, S. P., Liu, P., Chu, Y., Attie-Picker, M., Bahar Buyukbabani, M., Savulescu, J., Hannikainen, I. R., and Earp, B. D. (2025). Personalizing AI Art Boosts Credit, Not Beauty.

Ko, H.-K., Park, G., Jeon, H., Jo, J., Kim, J., and Seo, J. (2023). Large-scale Text-to-Image Generation Models for Visual Artists' Creative Works. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 919–933, Sydney NSW Australia. ACM.

Kohya-SS (2026a). musubi-tuner. GitHub repository. `https://github.com/kohya-ss/musubi-tuner`.

Kohya-SS (2026b). sd-scripts. GitHub repository. `https://github.com/kohya-ss/sd-scripts`.

Lee, B. C. and Chung, J. (2024). An empirical investigation of the impact of ChatGPT on creativity. *Nature Human Behaviour*, 8(10):1906–1914.

Li, M., Lin, Y., Zhang, Z., Cai, T., Li, X., Guo, J., Xie, E., Meng, C., Zhu, J.-Y., and Han, S. (2025). Svdquant: Absorbing outliers by low-rank components for 4-bit diffusion models.

Lian, L., Li, B., Yala, A., and Darrell, T. (2023). LLM-grounded Diffusion: Enhancing Prompt Understanding of Text-to-Image Diffusion Models with Large Language Models.

Liu, C., Takikawa, T., and Jacobson, A. (2024). A LoRA is Worth a Thousand Pictures.

Liu, V. and Chilton, L. B. (2022). Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *CHI Conference on Human Factors in Computing Systems*, pages 1–23. ACM.

Lovato, J., Zimmerman, J. W., Smith, I., Dodds, P., and Karson, J. L. (2024). Foregrounding Artist Opinions: A Survey Study on Transparency, Ownership, and Fairness in AI Generative Art. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7:905–916.

Meincke, L., Nave, G., and Terwiesch, C. (2025). ChatGPT decreases idea diversity in brainstorming. *Nature Human Behaviour*, 9(6):1107–1109.

Messer, U. (2024). Co-creating art with generative artificial intelligence: Implications for artworks and artists. *Computers in Human Behavior: Artificial Humans*, 2(1):100056.

Nichol, A. and Dhariwal, P. (2021). Improved Denoising Diffusion Probabilistic Models.

Peng, X. (2024). Designing Expressive Interaction with Generative Artificial Intelligence. In Lorig, F., Tucker, J., Dahlgren Lindström, A., Dignum, F., Murukannaiah, P., Theodorou, A., and Yolum, P., editors, *Frontiers in Artificial Intelligence and Applications*. IOS Press.

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., and Rombach, R. (2023). SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis.

Porquet, J., Wang, S., and Chilton, L. B. (2024). Copying style, Extracting value: Illustrators' Perception of AI Style Transfer and its Impact on Creative Labor.

Rajcic, N., Llano Rodriguez, M. T., and McCormack, J. (2024). Towards a Diffractive Analysis of Prompt-Based Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–15, Honolulu HI USA. ACM.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models.

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., and Aberman, K. (2022). DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation.

Shan, S., Cryan, J., Wenger, E., Zheng, H., Hanocka, R., and Zhao, B. Y. (2023). Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models.

Shi, J., Jain, R., Doh, H., Suzuki, R., and Ramani, K. (2023a). An HCI-Centric Survey and Taxonomy of Human-Generative-AI Interactions.

Shi, J., Jain, R., Duan, R., and Ramani, K. (2023b). Understanding Generative AI in Art: An Interview Study with Artists on G-AI from an HCI Perspective.

Sokhansanj, B. A. (2025). Local AI Governance: Addressing Model Safety and Policy Challenges Posed by Decentralized AI. *AI*, 6(7):159.

Unsloth (2026a). gemma-4-E2B-it-GGUF. Hugging Face model card. `https://huggingface.co/unsloth/gemma-4-E2B-it-GGUF`.

Unsloth (2026b). Unsloth Dynamic 2.0 GGUFs. Official documentation. `https://unsloth.ai/docs/basics/unsloth-dynamic-2.0-ggufs`.

Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q., and Velloso, E. (2024). The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18, Honolulu HI USA. ACM.

Wang, Z., Chen, C., Lyu, L., Metaxas, D. N., and Ma, S. (2023). DIAGNOSIS: Detecting Unauthorized Data Usages in Text-to-image Diffusion Models.

Wei, Y., Zheng, Y., Zhang, Y., Liu, M., Ji, Z., Zhang, L., and Zuo, W. (2025). Personalized image generation with deep generative models: A decade survey. *Computational Visual Media*, 11(6):1141–1194.

Wu, C., Yin, S., Qi, W., Wang, X., Tang, Z., and Duan, N. (2023). Visual ChatGPT: Talking, Drawing and Editing with Visual Foundation Models.

Xie, Y., Pan, Z., Ma, J., Luo Jie, and Mei, Q. (2023). A Prompt Log Analysis of Text-to-Image Generation Systems.

Yang, Z., Wang, J., Li, L., Lin, K., Lin, C.-C., Liu, Z., and Wang, L. (2023). Idea2Img: Iterative Self-Refinement with GPT-4v(ision) for Automatic Image Design and Generation.

Ye, H., Zhang, J., Liu, S., Han, X., and Yang, W. (2023). IP-adapter: Text compatible image prompt adapter for text-to-image diffusion models.

Yu, Y. (2025). A Review of Human-Centric Generative Image Editing: From Latent Space Control to Interactive Manipulation. *Applied and Computational Engineering*, 210(1):1–12.

Z-Image Team (2025). Z-Image: An Efficient Image Generation Foundation Model with Single-Stream Diffusion Transformer. Official repository. `https://github.com/Tongyi-MAI/Z-Image`.

Zeng, Q., Hu, C., Song, M., and Song, J. (2025). Diffusion Model Quantization: A Review.

Zhang, L., Rao, A., and Agrawala, M. (2023). Adding Conditional Control to Text-to-Image Diffusion Models.

Zhang, S. and Li, S. (2024). "Confrontation or Acceptance": Understanding Novice Visual Artists' Perception towards AI-assisted Art Creation.

Zhang, X., Wei, X., Hu, W., Wu, J., Wu, J., Zhang, W., Zhang, Z., Lei, Z., and Li, Q. (2025). A Survey on Personalized Content Synthesis with Diffusion Models. *Machine Intelligence Research*, 22(5):817–848.

Semi-Structured Interview Guide

Part 1: Usability & Workflow (Objective 1)

Goal: To understand the friction and usability of the interface.

1. **General Experience:** How would you describe the learning curve of the system you used today?
2. **Interface Friction:** Was there a specific moment where you felt lost or frustrated with the interface?

Part 2: Creative Utility (Objective 2)

Goal: To assess if the tool actually helped the creative process or just created distraction.

1. **Role of the Tool:** Did the generated images influence the direction of your sketch in a tangible way?
2. **Idea Generation:** Can you point to a specific element in your final sketch that came directly from an AI suggestion?
3. **Workflow Fit:** If you had this tool installed at what stage of your process would you use it? (e.g., Brainstorming, solving blockages, detailing, or never?)
4. **Creative Enhancement:** Do you feel that using the tool enhanced aspects of the creative process?

Part 3: Personalization & Fidelity (Objective 3)

Goal: To evaluate the success of the LoRA training and the fidelity of the workflow.

1. **Self-Recognition:** When you look at the AI outputs, do you recognize your own style in them?
2. **Success vs. Failure:** Were there specific aspects of your style that the model captured really well? Conversely, what did it consistently get wrong?

3. **Personalization benefits:** Was the personalization uniquely useful in some way?

Part 4: Control, Trust & Ethics (Objective 4)

Goal: To understand the value of the "Local/Private" paradigm.

1. **The Privacy Factor:** Knowing that this system runs entirely offline on a private server - and that your data is not being sent to a company - does that change how willing you are to use it?

2. **Ownership:** Who do you feel owns the images that were generated today? Would you view it differently if the tools were not using your previous works?

3. **Comparison:** How does this experience compare to your previous experiences with cloud-based AI tools?

Creative Task Selection

Instructions

Please select **ONE** of the following three task options for the upcoming 60-minute session.

Objective: To produce a preliminary concept sketch or study on your own device.

System Usage: The AI workstation is available to function as a visual assistant. You are encouraged to use it for reference generation, rapid iteration, and exploring stylistic possibilities or however else you want. You may transfer any generated images to your device.

OPTION A: Standardized Concept Briefs
Select this option if you prefer a structured starting point. Choose one of the following concepts to visualize in your own style.

---

1. **Character: "The Courier"**
   Design a character whose job is to transport important items across long distances. They should look equipped for travel and capable of handling a dangerous journey.
2. **Environment: "The Safehouse"**
   Design an exterior or interior shot of a hidden location used for shelter. The atmosphere should feel secure, isolated, and lived-in.
3. **Prop: "The Object"**
   Design an object of unknown origin that looks valuable or powerful.

OPTION B: Professional Work Integration
Select this option if you wish to apply the tool to an existing creative problem.

Context: Select a current project or personal artwork that you are currently developing or have found difficult to resolve.

Task: Utilize the session to advance this specific project.

OPTION C: Open-Ended Ideation
Select this option if you wish to explore the capabilities of the personalized model from a blank slate.

Context: You are starting from zero with no pre-conceived concept.

Task: Create something new from scratch.

Post-Session Questionnaire

Participant ID: _____ **Date:** _____

Instructions: Please indicate your level of agreement with the statements in the sections below. Remember: We are evaluating the software, not you. Your honest feedback is highly appreciated.

Part A: Usability

1 = Strongly Disagree 3 = Neutral 5 = Strongly Agree

Statement	**1**	**2**	**3**	**4**	**5**
1. I think that I would like to use this system frequently.	◯	◯	◯	◯	◯
2. I found the system unnecessarily complex.	◯	◯	◯	◯	◯
3. I thought the system was easy to use.	◯	◯	◯	◯	◯
4. I think that I would need the support of a technical person to be able to use this system.	◯	◯	◯	◯	◯
5. I found the various functions in this system were well integrated.	◯	◯	◯	◯	◯
6. I thought there was too much inconsistency in this system.	◯	◯	◯	◯	◯
7. I would imagine that most people would learn to use this system very quickly.	◯	◯	◯	◯	◯
8. I found the system very cumbersome to use.	◯	◯	◯	◯	◯
9. I felt very confident using the system.	◯	◯	◯	◯	◯
10. I needed to learn a lot of things before I could get going with this system.	◯	◯	◯	◯	◯

Part B: Creative Utility

1 = Strongly Disagree 3 = Neutral 5 = Strongly Agree

Statement	**1**	**2**	**3**	**4**	**5**
1. The generated images helped me generate new ideas or concepts.	◯	◯	◯	◯	◯
2. The images were useful as visual references for my own drawing.	◯	◯	◯	◯	◯
3. I felt more productive using this tool alongside my main work.	◯	◯	◯	◯	◯
4. The quality of the images was high enough to be useful.	◯	◯	◯	◯	◯
5. I would use a tool like this in my professional studio workflow.	◯	◯	◯	◯	◯

Part C: Personalization & Style

1 = Strongly Disagree 3 = Neutral 5 = Strongly Agree

Statement	**1**	**2**	**3**	**4**	**5**
1. The AI outputs accurately reflected my personal artistic style.	◯	◯	◯	◯	◯
2. The generated images felt "authentic" to my brand/identity.	◯	◯	◯	◯	◯
3. The model captured the specific textures and details I use.	◯	◯	◯	◯	◯
4. I preferred these personalized results over generic AI images.	◯	◯	◯	◯	◯

Part D: Control, Privacy & Trust

1 = Strongly Disagree 3 = Neutral 5 = Strongly Agree

Statement	**1**	**2**	**3**	**4**	**5**
1. I felt in control of the creative output.	◯	◯	◯	◯	◯
2. Knowing this system was private/local increased my willingness to use it.	◯	◯	◯	◯	◯
3. I felt a sense of ownership over the images generated today.	◯	◯	◯	◯	◯
4. I would feel comfortable using this system for confidential or unreleased professional work.	◯	◯	◯	◯	◯

Part E: Final Thoughts

1. What was the biggest barrier to using this tool today?

2. If you could change one thing about the workflow, what would it be?

3. Any other comments?