Explainability-in-Action: Enabling Expressive Manipulation and Tacit Understanding by Bending Diffusion Models in ComfyUI

Ahmed M. Abuzuraiq
School of Interactive Arts and Technology
Simon Fraser University
Surrey, Canada
aabuzura@sfu.ca

Philippe Pasquier
School of Interactive Arts and Technology
Simon Fraser University
Surrey, Canada
pasquier@sfu.ca

Abstract

Explainable AI (XAI) in creative contexts can go beyond transparency to support artistic engagement, modifiability, and sustained practice. While curated datasets and training human-scale models can offer artists greater agency and control, large-scale generative models like text-to-image diffusion systems often obscure these possibilities. We suggest that even large models can be treated as creative materials if their internal structure is exposed and manipulable. We propose a craft-based approach to explainability rooted in long-term, hands-on engagement—akin to Schön's "reflection-in-action"—and demonstrate its application through a model-bending and inspection plugin integrated into the node-based interface of ComfyUI. We demonstrate that by interactively manipulating different parts of a generative model, artists can develop an intuition about how each component influences the output.

Keywords

Explainable AI, Generative AI, Diffusion Models, Visual Art, Model Bending, Model Crafting, Reflection-in-Action

1 Introduction

Explainable AI (XAI) traditionally focuses on demystifying machine learning systems for auditing, transparency, or safety [16]. However, explainability in creative contexts can serve different roles, including making models modifiable and debuggable for a meaningful artistic engagement [9], and creating the conditions for a sustained artistic practice that goes beyond mere amusement [27]. The last workshop on XAIxArts [10] culminated in a manifesto that called artists and researchers to explore alternatives to the technocentric explanations of AI that value artistic practices and use

intentional hacking, glitches, and imperfections as creative tools. This work continues along those lines.

It is argued that working with curated (small) datasets and human-scale models can enhance artists' agency [29]. Broadening the scope of the artistic process with AI to include both model training and inference can also reinforce artists' trust in AI models [27]. Both approaches afford artists more control over AI models. However, artists' ability to craft with AI models as materials for creation [6, 11, 24] is diminished when working with large-scale generative models [4]. If small-data and model training present an alternate explainable way for artists working with human-scale generative models, what options are available for artists wanting to work with large-scale models such as text-to-image diffusion models? Although this material relationship is easier to build with small models, where data and model structure/training are accessible and malleable, it becomes more challenging at scale. However, we argue that large models can also be treated as material, provided their structure is exposed and can be manipulated.

We propose that treating AI models as materials, especially within interfaces that expose the model's components like the node-based interface of ComfyUI, can foster a craft-based relationship between artists and generative systems. In particular, through long-term engagement and hands-on manipulation of a model's components, artists can develop a form of tacit understanding and familiarity with their AI materials akin to that found in traditional crafts, i.e. a form of explainability that is rooted in doing, aligned with Schön's "reflection-in-action" [25]. We explore the application of these ideas for large-scale models through a plugin for model bending [8] that is implemented into the node-based interface of ComfyUI [12].

2 From Models as Commodity to Models as Material

Model repositories like CivitAI [1] and TensorArt [2] provide users with thousands of pre-trained and fine-tuned models, generated images with associated prompts and parameters, and reproducible workflows for a wide variety of use-cases and styles. The quantity and diversity of models and workflows shared online are simply unprecedented within the context of generative art, and cloud computing services are making them more accessible. A large portion of these models are adaptations and personalizations (e.g. Low-Rank Adaptations – LoRAs [19]) of a smaller set of base models (e.g. Stable Diffusion [26] and Flux variants). Following Abonamah et al. [3], we argue that the proliferation of generative AI models will likely result in their commodification — a development that, we contend, may further perpetuate their biases [28].

When technology becomes commodified, its design priorities shift—from serving as a raw material for creation (as with early computer terminals) to emphasizing accessibility and replaceability. This shift can bring innovation benefits, as seen with personal computers. However, in the case of generative AI, commodification is occurring at an unprecedented pace. The challenge, however, is that this rapid commodification may discourage sustained engagement with individual models. Instead of fostering deep familiarity with their inner workings, affordances, and limitations, users are incentivized to sample from an ever-expanding catalogue without developing the tacit knowledge necessary for critical or innovative use. This dynamic may also contribute to *AI fatigue*—a feeling of exhaustion or overwhelm, potentially driven by the rapid pace of updates to AI models and tools, as well as the opaqueness and complexity of AI systems [1]. Historically, artists have cultivated a deep, material understanding of their tools—whether brushes, cameras, or software—as a prerequisite for meaningful creative expression. Generative models require the same kind of sustained engagement to be used critically and responsibly. At the same time, many artists—who are uniquely positioned to critique and subvert these systems [6, 10]—are choosing not to engage with them on principle, citing concerns that these models are trained on the work of unattributed artists [20]. Therefore, we argue that it's imperative to encourage and offer tools for artists to engage with AI in ethical [2], sustained [27] and critical [6] ways, and explainability facilitates these forms of engagement.

3 Explainability for Art Practice vs. Art Practice for Explainability

We draw a distinction between two complementary roles of explainability in the context of AI and art. First, explainability can support artists working with AI by enhancing control, agency, and trust [27]. Second, artistic practice can, in turn, contribute to broader societal and cultural understanding of AI—e.g., by raising questions about its production and use [6, 17]. Explainability can support a behavioural understanding of AI models, i.e. a mental model of how inputs map to outputs. In this work, we argue that explainability is valuable insofar as it helps artists achieve their desired outcomes [5], rather than providing a comprehensive account of the model's inner workings. However, sustained and effective artistic engagement with generative AI often requires some degree of theoretical understanding, such as recognizing the roles of different parameters and components within a generative system, and how these can be manipulated to produce novel outputs that extend beyond the model's original training or design [7]. As we will show next, art creation tools that expose and allow the manipulation of different components within a generative AI system can foster both forms of understanding.

4 Case Study: Diffusion Model Bending and Inspection in ComfyUI

ComfyUI is an open-source, node-based interface for designing and executing advanced diffusion pipelines. Unlike other interfaces such as AUTOMATIC1111 [30], which prioritizes simplicity and abstracts away the model's internal workings, ComfyUI adopts a low-code, modular approach that decomposes the diffusion process into discrete nodes. These nodes represent stages such as loading base models and applying LoRA adaptations, encoding images into latent representations, generating textual embeddings via the CLIP model, and configuring the denoising process in diffusion. ComfyUI's backend also manages model loading in a way that helps circumvent GPU memory constraints.

We chose ComfyUI because it exposes the inner components of the diffusion pipeline, encouraging users to explore, customize, and develop an understanding of each component. Additionally, ComfyUI frequently integrates state-of-the-art image generation models, making it a timely and flexible testbed for experimentation. A basic workflow in ComfyUI is illustrated in Figure 1, and includes steps such as model loading, prompt encoding with CLIP, sampling with user-defined parameters (e.g., number of steps, CFG scale), and decoding the resulting latent image.

4.1 Related Work

Network bending is inspired by the tradition of circuit bending—a practice originating in experimental electronic music, in which artists intentionally short-circuit or rewire electronic devices (often children's toys or synthesizers) to produce unpredictable or expressive behaviours. In the context of machine learning, network bending refers to the deliberate manipulation of a model's internal activations or computational pathways to intervene in its generative process, typically for expressive or exploratory purposes [8]. This is achieved by injecting a *bending operator* into a model to transform its intermediate outputs during inference, without the need for additional training or data. Bending operators can perform operations such as adding noise, multiplying/adding scalar values, rotation, and scaling; morphological operations like erosion and dilation, or other custom operations. Previous work on network bending has applied interventions at various layers of the synthesizer network in StyleGAN models [8, 21, 23]. In the context of diffusion models, bending operations have been implemented at user-specified timesteps within the denoising process [14] (as inferred from source code), and at select layers of the noise prediction model (UNet) [13].

4.2 System Design

In this work, we implement a model bending system as a suite of custom nodes into the ComfyUI interface. The overarching goals of this work are: (1) to provide tools for introducing variations and diversity into the text-to-image diffusion generative process, and (2) these tools should also facilitate a better understanding of the generative process and its parts. The bending operations we implement interject the flow in Figure 1 by processing the UNet/VAE/LoRA/text-embeddings before passing them downstream.

In particular, we introduce a set of tools within ComfyUI that enable artists to bend various components of the latent diffusion pipeline, including but not limited to the components covered by previous work [8, 13, 14]. These components include the Variational Autoencoder (VAE) used to encode and decode images, the

---

[1]https://newsletter.victordibia.com/p/you-have-ai-fatigue-thats-why-you
[2]The CommonCanvas [15], trained on Creative Commons images, may offer an alternative for artists who find using large generative models questionable.
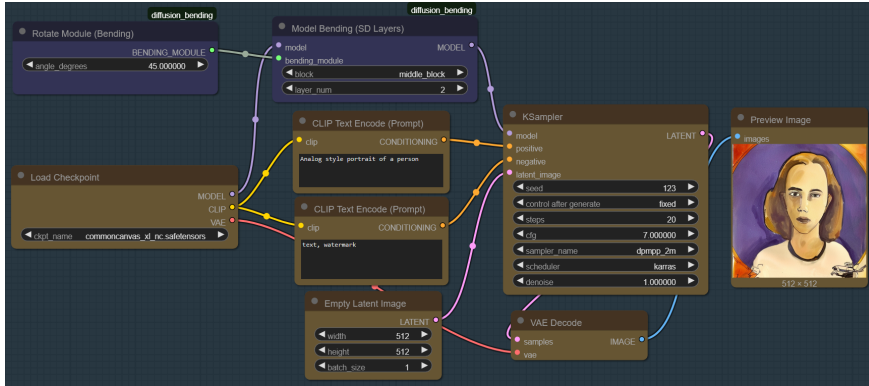
Figure 1: Basic workflow in ComfyUI (gold) running on the open Common Canvas model [15]. The UNet bending nodes shown in purple.
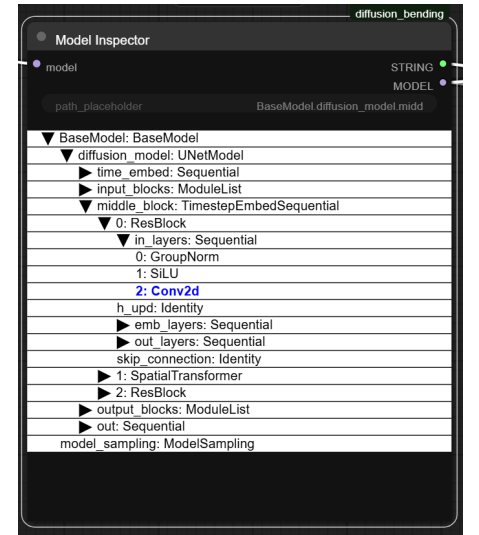

Figure 2: With the model inspector node, artists can interactively pick layers in a model to manipulate. Picked layers are highlighted.

text embeddings produced by the Contrastive Language-Image Pre-training (CLIP) model, the noise prediction process handled by the UNet, as well as at select timesteps during the sampling process. Bending can be used to modify outputs at specific points in the latent diffusion pipeline and to introduce controlled variations, particularly when other generation parameters remain fixed. Our use of the term bending aligns with the ethos of Broad et al.[8], i.e., applying transformations to specific parts of a generative process or model for creative and exploratory purposes. However, it also differs from their original definition in key ways. In our implementation, bending does not exclusively imply injecting custom operators into the model, nor does it require the explicit extraction and manipulation of semantic features. This is in part because large-scale text-to-image diffusion models appear to exhibit semantically rich latent spaces [3].

The tool can be downloaded manually through a git repository [4] or by installing it through the Comfy Registry [5]. For brevity, we describe essential parts of the system next and refer those interested to the project's website for details and sample outputs.

4.3 UNet Model Bending

Denoising diffusion models recreate images by first adding noise to them in steps, and then learning how to undo that noise. During training, a UNet model learns to clean the noise, so it can later produce new images during inference starting from random inputs [18]. Given their central role in image generation, UNets are a prime target for bending. UNet models form a U shape with gradually shrinking input blocks that feed to middle blocks, which then expand again to recreate the input images through output blocks.

---

[3]Though certain parts (*h-space*) seem to be more potent for semantic manipulation[22]
[4]https://github.com/abuzreq/ComfyUI-Model-Bending
[5]https://registry.comfy.org/publishers/abuzreq/nodes/ComfyUI-Model-Bending

To bend a UNet model, the system requires two components: a bending operator (predefined or custom operators) and a path to the layer to be bent (e.g., *diffusion_model.middle_block.0.in_layers*). For specifying the path to the layer to be bending, users can use the *Model Bending (SD Layers)* node, pick the blocks within the UNet to bend, and a convolution layer within that block (Figure 1). Once inserted, the bending module becomes part of the model's inference pipeline, and it is invoked during each denoising step that follows.

4.4 Model Inspection

For more granular control—including the ability to select specific blocks or layers—we provide the Model Inspector node (Figure 2). This tool displays the model's architecture as a nested, expandable tree, allowing users to visually navigate and select any layer to pass to the model bending node. It currently supports both UNet and VAE models. By interactively bending different parts of the model, artists can develop intuition about how each component influences the output. Figure 3 illustrates bending with a rotation operation for select layers in the UNet of the Stable Diffusion v1.4 model.

4.5 Feature Maps Visualization

Feature map visualizations are often used to understand the model's behaviour and the kind of features it is learning at different layers. We provide a Visualize Feature Map node, which allows users to visualize intermediate feature maps at any layer, and the Model Inspector can be used to pick the layers to be visualized interactively. By selecting a layer of interest, artists can inspect what the model is attending to at different stages of the denoising process. These visualizations not only help reveal the evolution of the output but can also be paired with the Model Bending node to observe how specific modifications affect the model's internal representations.
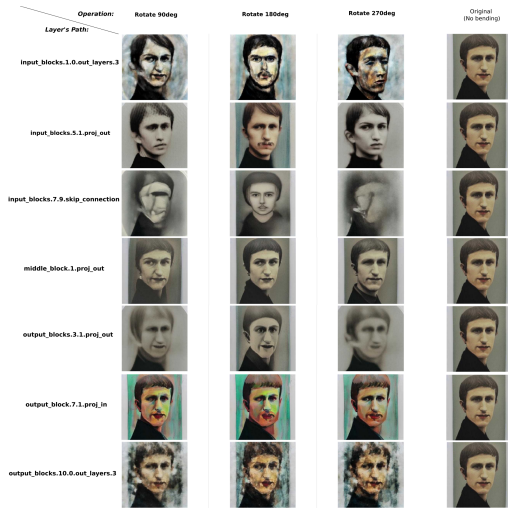
Figure 3: A rotation bending applied to select layers of the SDv1.4 model, prompt "analog style portrait of a person", seed 42, 20 steps, CFG 7, dpmpp_2m sampler, and Karras scheduler. The result with no bending is shown on the right.

4.6 CLIP Text Encoding (Conditionings)

In a text-to-image system, prompts are encoded into textual embeddings using the CLIP model (referred to as *conditioning*). Our plugin enables users to make fine-grained adjustments within the embedding space, specifically in the local region where the prompt is mapped. Such small movements within the embedding space can help artists understand its structure and offer fine controls that complement prompt-based interaction.

5 Conclusion and Future Work

By reframing explainability as an artistic practice grounded in making and manipulation, this paper offers a different lens for thinking about AI-human collaboration. The model bending plugin for ComfyUI invites artists to engage with the inner workings of the large-scale models themselves, treating them as expressive, deformable materials. In the future, we plan to extend this tool with additional forms of model manipulation and custom semantic editing. We also aim to build a community of practice around this tool. Artists might share recipes or suggestions for which parts of a large model yield the most expressive outcomes when bent or their critical reflections while using the tool to inspect models. We hope this approach fosters a deeper, more personal relationship between artists and AI models—one where explainability emerges not just from observation, but from doing. Ultimately, we believe that model crafting tools like the one presented here, which encourage artists to treat large-scale generative models as creative materials, may help address ongoing challenges around authorship and agency in GenAI-assisted art.

References

[1] 2025. Civitai: The Home of Open-Source Generative AI. https://civitai.green/.
[2] 2025. Tensor.art: AI model sharing platform. https://tensor.art/.
[3] Abdullah A Abonamah, Muhammad Usman Tariq, and Samar Shilbayeh. 2021. On the commoditization of artificial intelligence. *Frontiers in psychology* 12 (2021).
[4] Ahmed M. Abuzuraiq and Philippe Pasquier. 2024. Seizing the Means of Production: Exploring the Landscape of Crafting, Adapting and Navigating Generative AI Models. In *the 3rd Generative AI and HCI Workshop* (Hawaii, USA).
[5] Aditya Bhattacharya. 2024. Towards Directive Explanations: Crafting Explainable AI Systems for Actionable Human-AI Interactions. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems.* 1–6.
[6] Terence Broad. 2024. Using Generative AI as an Artistic Material: A Hacker's Guide. In *In Proceedings of Explainable AI for the Arts Workshop 2024.*
[7] Terence Broad, Sebastian Berns, Simon Colton, and Mick Grierson. 2021. Active Divergence with Generative Deep Learning–A Survey and Taxonomy. In *Proceedings of the 12th International Conference on Computational Creativity (ICCC '21).* doi:arXivpreprintarXiv:2107.05599
[8] Terence Broad, Frederic Fol Leymarie, and Mick Grierson. 2021. Network Bending: Expressive Manipulation of Deep Generative Models. In *Artificial Intelligence in Music, Sound, Art and Design: 10th International Conference, EvoMUSART 2021, Held as Part of EvoStar 2021, Virtual Event, April 7–9, 2021, Proceedings 10.* Springer.
[9] Nick Bryan-Kinns, Berker Banar, Corey Ford, Courtney N Reed, Yixiao Zhang, Simon Colton, and Jack Armitage. 2023. Exploring XAI for the Arts: Explaining Latent Space in Generative Music. (2023). arXiv:2308.05496
[10] Nick Bryan-Kinns, Shuoyang Jasper Zheng, Francisco Castro, Makayla Lewis, Jia-Rey Chang, Gabriel Vigliensoni, Terence Broad, Michael Clemens, and Elizabeth Wilson. 2025. XAIxArts Manifesto: Explainable AI for the Arts. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan, 2025). doi:10.1145/3706599.3716227 arXiv:2502.21220 [cs]
[11] Baptiste Caramiaux and Sarah Fdili Alaoui. 2022. "Explorers of Unknown Planets" Practices and Politics of Artificial Intelligence in Visual Arts. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.
[12] comfyanonymous. 2024. https://github.com/comfyanonymous/ComfyUI.
[13] Garin Curtis. 2025. *Beyond Prompts: Developing An Expressive Tool for Real-Time Manipulation of Diffusion Models through Active Divergence.* https://garincurtis.com/projects/beyond-prompts
[14] Luke Dzwonczyk, Carmine Emanuele Cella, and David Ban. 2024-06-28. Network Bending of Diffusion Models for Audio-Visual Generation. arXiv. doi:10.48550/arXiv.2406.19589 arXiv:2406.19589 [cs]
[15] Aaron Gokaslan, A Feder Cooper, Jasmine Collins, Landan Seguin, Austin Jacobson, Mihir Patel, Jonathan Frankle, Cory Stephenson, and Volodymyr Kuleshov. 2024. CommonCanvas: Open diffusion models trained on creative-commons images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 8250–8260.
[16] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. *Science robotics* 4, 37 (2019).
[17] Drew Hemment, Dave Murray-Rust, Vaishak Belle, Ruth Aylett, Matjaz Vidmar, and Frank Broz. 2024. Experiential AI: Between arts and explainable AI. *Leonardo* 57, 3 (2024), 298–306.
[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. 33 (2020), 6840–6851.
[19] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank Adaptation of Large Language Models. (2021). arXiv:2106.09685
[20] Reishiro Kawakami and Sukrit Venkatagiri. 2024. The Impact of Generative AI on Artists. In *Proceedings of the 16th Conference on Creativity & Cognition.* 79–82.
[21] Jonas Kraasch and Philippe Pasquier. 2022. Autolume-Live: Turning GANs into a Live VJing Tool. In *Proceedings of the 10th Conference on Computation, Communication, Aesthetics & X* (Coimbra, Portugal). 152–169.
[22] Mingi Kwon, Jaeseok Jeong, and Youngjung Uh. 2023. *Diffusion Models Already Have a Semantic Latent Space.* doi:10.48550/arXiv.2210.10960 arXiv:2210.10960 [cs]
[23] Metacreation Lab. 2024. Autolume: A Neural-network Based Visual Synthesizer. https://www.metacreation.net/autolume.
[24] Sonja Rozental, Michel Van Dartel, and Alwin De Rooij. 2025. How Artists Use AI as a Responsive Material for Art Creation. doi:10.31234/osf.io/gjdnw
[25] Donald A Schön. 1983. *The Reflective Practitioner: How Professionals Think in Action.* New York: Basic Books.
[26] StabilityAI. 2024. Stable Diffusion. https://stability.ai/stable-image.
[27] Austin Tecks, Thomas Peschlow, and Gabriel Vigliensoni. 2024. Explainability Paths for Sustained Artistic Practice with AI. In *In Proceedings of Explainable AI for the Arts Workshop 2024 (XAIxArts 2024)* (2024-07-21). arXiv. doi:10.48550/arXiv.2407.15216 arXiv:2407.15216 [cs]
[28] Adriana Fernández de Caleya Vázquez and Eduardo C Garrido-Merchán. 2024. A Taxonomy of the Biases of the Images created by Generative Artificial Intelligence. *Current Trends in Business Management 3* (2024).
[29] Gabriel Vigliensoni, Phoenix Perry, and Rebecca Fiebrink. 2022. A Small-Data Mindset for Generative AI Creative Work. In *In Generative AI in HCI Workshop, CHI '22* (New York, NY, USA). 5.
[30] Automatic1111 WebUI. 2024. https://github.com/AUTOMATIC1111/stable-diffusion-webui.