

From Code to Camera: The Making and Meaning of Prosomoíosi (Simulation), an AI Documentary Film

Anonymous Author(s)

Anonymous Author(s)

Anonymous Institution

Anonymous City, Anonymous Country

anonymous@email.address

pics Prosomoiosi 4pages short paper/band.jpg

Figure 1: Screenshot of Prosomoiosi (Simulation), generated from Cultural Analytics's cover[?]

Abstract

This paper investigates the audiovisual artwork “Prosomoiosi (Simulation)”, advancing the notion of “medium alignment” to critique how successive simulation technologies overwrite cultural memory. Grounded in media-archaeological thinking, the study argues that concepts must be articulated through media whose operative logic remains visible, thereby transforming viewers from passive

spectators into active witnesses of algorithmic decision-making. A live diffusion pipeline—disclosed rather than concealed—functions as both method and message, allowing audiences to see how text prompts, stochastic noise, and performer input co-evolve on screen. By tracing precedents from early interactive installations to recent AI-driven pieces, the paper situates Prosomoiosi within a lineage that questions tool-centric spectacle and re-negotiates authorship

at the human-machine frontier. Medium alignment thus emerges as a transferable design heuristic for artists seeking to move beyond technological virtuosity toward works that openly expose, rather than obscure, the politics of their own making.

CCS Concepts

• **Applied computing** → **Media arts**; • **Computing methodologies** → *Artificial intelligence*.

Keywords

generative AI, medium alignment, real-time diffusion, text-to-video synthesis, StreamDiffusion, TouchDesigner, AI film, media archaeology, fiction documentary

1 Introduction

Over the past decade, the impact of generative artificial intelligence (AI) on audiovisual media has moved rapidly from research laboratories to the core workflows of the film industry. In 2024 the Academy of Motion Picture Arts and Sciences explicitly allowed films containing AI-generated footage to compete for Oscars—a symbolic step that not only reflects Hollywood’s recognition of a technological paradigm shift but also signals a loosening of traditional production standards. In the same year, the Runway AI Film Festival (New York and Los Angeles) and the RAIN Film Fest (Barcelona) were launched, indicating that AI-based image creation is transitioning from peripheral artistic practice to a scaled ecosystem; the inauguration of the first Asian AI Film Festival (HKUST AI Film Festival) in 2025 further confirms the global spread and cultural diversification of this trend.

Technically, breakthroughs in generative models for script writing, storyboarding, image synthesis, and even long-form video generation have opened unprecedented real-time iterative spaces for creators. Text-to-video models such as Runway Gen-2¹, OpenAI’s Sora², and Google’s Veo³, built on large-scale spatio-temporal diffusion frameworks, produce controllable, high-definition, and narratively coherent output; meanwhile, text-to-image models like Stable Diffusion⁴, Flux⁵, and Midjourney⁶, together with large language models (LLMs) for script development and world-building, are reshaping the traditional pre-production workflow. Researchers note that AI tools not only increase production efficiency but also introduce a “machine vision” cultural grammar that co-evolves with audience perception.

Yet this technological surge exposes a largely overlooked contradiction: when artists embrace a new medium while retaining an old conceptual framework—or conversely, use new concepts to drive an old medium—a misalignment arises between medium and idea. The consequences are typically:

- (1) The new technology is treated merely as an “accelerator,” unable to participate substantively in the work’s intent;

- (2) The concept is forced to yield to the tool’s default paradigm, resulting in demonstrative or techno-virtuosic superficiality;
- (3) Audiences perceive chiefly a “technological spectacle” rather than the work’s thematic core, thereby weakening attention to the artistic statement itself.

Generative AI is rapidly reshaping film and art production, yet conceptual thinking has not kept pace, widening the gap between medium and idea. Using the video work *Prosomoiosi* (Simulation) as a case study, we advance a speculative notion of “medium alignment”: aligning a project’s theme with the operative logic of its medium from the outset so that the algorithmic pipeline becomes a visible narrative layer rather than backstage infrastructure. This proposal echoes Baudrillard’s model-first simulation logic [?], Bolter and Grusin’s remediation paradigm [?], and Manovich’s analysis of algorithmic storytelling [?]. *Prosomoiosi* literalises “simulation”: real-time images are produced in TouchDesigner⁷ via StreamDiffusion⁸, driven by key-framed parameters, then tiled and up-scaled to 4K with ComfyUI and Flux model. By exposing the diffusion process on-screen, the work lets viewers watch algorithms overwrite the image itself. We argue that such medium alignment enables generative video to transcend tool fetishism and techno-virtuosity, turning it into a dynamic arena where memory, history, and subjectivity can be renegotiated.

2 Reference Artworks

To contextualize our research, this section examines a selection of artworks that trace the evolution from early interactive systems to contemporary AI-driven critiques of perception. A foundational example is *A-Volve* (1994) by Christa Sommerer and Laurent Mignonneau, a pioneering installation where visitors’ drawings are instantiated as virtual creatures [?]. The work foregrounds themes of artificial life and embodied interaction, positioning the human as a direct creator within a digital ecosystem. Decades later, this focus on perception and digital life has been radically reconfigured by deep learning. Rather than granting users direct creative agency, many contemporary artists use AI to interrogate the very process of seeing. For instance, Memo Akten’s *Learning to See* (2017) employs real-time neural networks to reveal how a model reconstructs reality through the biased filter of its training data, explicitly questioning the relationship between perception and truth [?]. A related, though aesthetically distinct, exploration is Scott Eaton’s *Entangled II* (2019), which uses a bespoke neural network to process abstract footage into morphing, quasi-human figures, examining the aesthetics of machine perception and the human tendency for pareidolia [?]. Taking this critique of data’s influence a step further, *POSTcard Landscapes from Lanzarote* (2022) by Varvara & Mar leverages a StyleGAN2 model trained on distinct datasets to generate two opposing visual realities of the same location: the tourist’s view and the local’s view [?]. By juxtaposing these AI-generated “gazes,” the work offers a powerful commentary on how curated data not only reflects but actively shapes cultural identity and memory, echoing John Urry’s concept of the “tourist gaze” [?]. The author of the present paper likewise extended this trajectory with

¹<http://runwayml.com>

²<https://openai.com/sora/>

³<https://deepmind.google/models/veo/>


⁴<https://huggingface.co/CompVis/stable-diffusion-v1-4>

⁵<https://huggingface.co/black-forest-labs/FLUX.1-dev>

⁶<https://www.midjourney.com/>

⁷<https://derivative.ca/>

⁸<https://github.com/cumulo-autumn/StreamDiffusion>



pics Prosomoiosi 4pages short paper/compare.jpg

Figure 2: Image quality comparison, left:previous work; right: this work

a 2024 artwork that employs real-time generative techniques to explore a symbiotic relationship between AI and human creators, the quality of generated videos comparison is below in Figure 2. Together, these works illustrate a significant shift from celebrating direct interaction with artificial life to a critical investigation of how AI models, and the data they are fed, mediate and construct our reality.

3 Case Study : Prosomoiosi (Simulation)

3.1 Artwork Description

Prosomoiosi (Simulation) is a video work created through a real-time generative pipeline that, through a media-archaeological lens, advances the notion of "medium alignment" to probe how successive technological regimes continually overwrite memory and

identity. The work pursues two mutually reinforcing aims: (1) to unmask, through a stratified narrative of simulation logics, the quiet manner in which each new medium overprints cultural remembrance; and (2) to disclose the algorithmic selectivity and symbolic authority of AI-driven image synthesis by staging a reflexive demonstration built on a TouchDesigner–StreamDiffusion pipeline featuring real-time multiprompt editing and ComfyUI up-scaling with the flux-dev f16 model. By allowing text prompts, stochastic noise, and live performer input to co-evolve on the exhibition floor, the piece shifts image production from depiction to ontological negotiation, inviting audiences to rethink the future interplay of human creativity and machine simulation.

Simulation here is not mere copying or representation; rather, it is a power-laden rewriting produced through the interplay of technology, archives, and algorithms. Wolfgang Ernst notes that archival temporality shapes memory even as it quietly edits the past [?]; Walter Benjamin anticipated the aura's erosion under mechanical reproduction [?], and Jean Baudrillard later warned of simulacra supplanting reality [?]. Charting a trajectory "from clay to code"—from sand tables and armillary spheres to deep learning and GANs—modeling technologies have evolved from cognitive tools to arbiters of truth: large networks such as GPT-4 and ESM-2 autonomously conjure worlds from latent space and even guide human decision-making, reshaping the human–reality relation. Within this landscape, video games have become the most pervasive simulation medium; by breaching the "fourth wall," they place players in a hybrid third space where procedural rhetoric lets symbolic capital and ideology permeate interaction, extending McLuhan's dictum that "the medium is the message" [?] alongside Bourdieu's analysis of symbolic power [?]. Educational and political simulations—SimCity, PeaceMaker—as well as Lorenz-style chaotic systems further attest to simulation's profound influence on behaviour and cognition. Faced with the medium mutations of algorithmic culture, art must enact medium alignment, expanding AI alignment into a perceptual-symbolic calibration at the level of the medium itself; situated between Stiegler's "third memory" [?] and Yuk Hui's "cosmic technics," [?] this posture seeks to cultivate a future symbiosis among humans, machines, and media.

3.2 Production Workflow

The real-time generation pipeline of this project consists of four stages: 'material preparation → diffusion generation → high-resolution upscaling → post-production,' with each stage optimized for live performance.

Material Preparation. Source footage includes public-domain vintage films and live gameplay recordings, uniformly transcoded to 24 fps H.264 in Adobe Premiere Pro. The H.264 codec utilizes advanced video compression techniques including motion estimation, transform coding, and entropy coding to achieve optimal file sizes while maintaining visual quality⁹. To optimize diffusion processing on RTX 4090 hardware, the footage is slowed to 4-14 fps (0.2×-0.5× speed) and streamed through NDI (Network Device Interface)¹⁰ to TouchDesigner for real-time generation.

The frame rate reduction strategy is crucial for maintaining real-time performance as diffusion models have significant computational overhead. Each frame requires multiple denoising steps through the U-Net architecture¹¹, making higher frame rates computationally prohibitive on current hardware. The NDI protocol enables low-latency, high-quality video streaming over standard Ethernet networks, providing crucial timing precision for real-time generative workflows¹².

Real-Time Diffusion Generation (TouchDesigner + StreamDiffusion plug-in). Hardware platform: AMD Ryzen 9 7950X / NVIDIA RTX 4090 (24 GB VRAM). In TouchDesigner, the StreamDiffusion node developed by dotsimulate¹³ is loaded. StreamDiffusion is an optimization framework that enables near real-time diffusion model inference by implementing stream batching and residual classifier-free guidance¹⁴. After configuring the project's dependencies, every video frame undergoes image-to-image (img2img) diffusion using denoising diffusion probabilistic models (DDPMs)¹⁵.

Models such as Stable Diffusion XL¹⁶, Stable Diffusion 1.5¹⁷, and their LoRAs (Low-Rank Adaptations)¹⁸ are switched dynamically to achieve desired artistic effects. The latent diffusion architecture operates in a compressed latent space using a variational autoencoder (VAE), significantly reducing computational requirements compared to pixel-space diffusion¹⁹.

StreamDiffusion parameters—Prompt, CFG Scale (Classifier-Free Guidance)²⁰, Steps, Denoise, etc.—are bound to keyframe curves via the Animation COMP curve editor in TouchDesigner, allowing precise control over all generation parameters. Performers can pre-program or live-edit these curves to sculpt narrative rhythm, achieving smoother and more coherent results through temporal consistency optimization.

The workflow also integrates ControlNet's HED (Holistically Nested Edge Detection)²¹ branch within TouchDesigner; by parameterizing the edge-strength weight, the diffusion process is further constrained to converge on scene geometry and spatial alignment. ControlNet enables fine-grained control over the generation process by conditioning the diffusion model on additional input modalities such as edge maps, depth maps, or pose estimations²².

¹¹Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241).

¹²Perkins, C. (2003). RTP: Audio and Video for the Internet. Addison-Wesley Professional.

¹³<https://dotsimulate.com/>

¹⁴Yamamoto, A., et al. (2023). StreamDiffusion: A Pipeline-level Solution for Real-time Interactive Generation. arXiv preprint arXiv:2312.12491.

¹⁵Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.

¹⁶Podell, D., et al. (2023). SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. arXiv preprint arXiv:2307.01952.

¹⁷Rombach, R., et al. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695).

¹⁸Hu, E. J., et al. (2021). LoRA: Low-Rank Adaptation of Large Language Models. arXiv preprint arXiv:2106.09685.

¹⁹Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

²⁰Ho, J., & Salimans, T. (2022). Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598.

²¹Xie, S., & Tu, Z. (2015). Holistically-nested edge detection. In Proceedings of the IEEE international conference on computer vision (pp. 1395-1403).

²²Zhang, L., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. arXiv preprint arXiv:2302.05543.

⁹Wiegand, T., et al. (2003). Overview of the H. 264/AVC video coding standard. IEEE Transactions on circuits and systems for video technology, 13(7), 560-576.

¹⁰NewTek NDI SDK Documentation: <https://www.ndi.tv/sdk/>

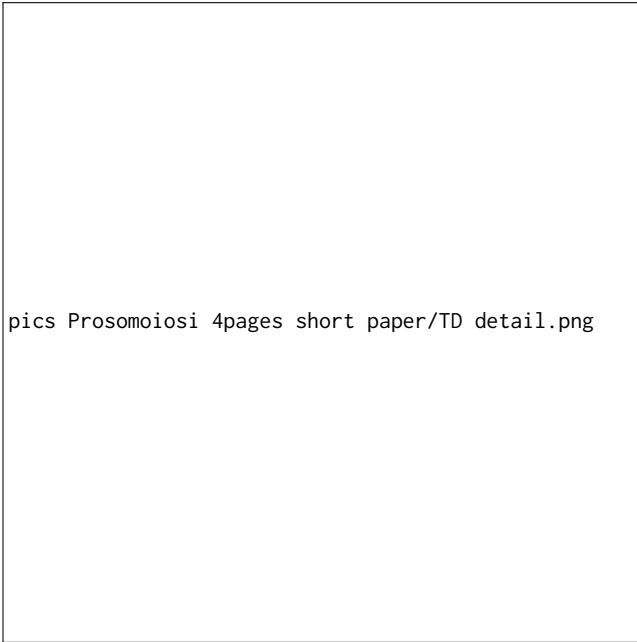


Figure 3: Screen capture of the TouchDesigner workflow: red – Animation COMP-related modules; green – StreamDiffusion node; yellow – ControlNet

High-Resolution Upscaling (ComfyUI Flux Workflow). The Flux pipeline first loads each image, resizes it to 1024×1024 for VRAM-adaptive processing, then applies a $4\times$ NMKD-SiAx super-resolution pass; this tile-based routine balances memory, accommodates large canvases, and boosts texture and edge detail for later stages. Flux represents a new generation of diffusion models that utilizes flow matching instead of traditional denoising approaches²³, enabling more efficient training and inference while maintaining high-quality output.

The NMKD-SiAx upscaling model is based on Real-ESRGAN architecture²⁴, which employs a generative adversarial network (GAN) specifically designed for real-world image super-resolution. The model utilizes Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN)²⁵ with improved perceptual loss functions and residual-in-residual dense blocks for better texture preservation.

To reduce VRAM load at high resolutions, we use the Tile-to-Patch (TTP) toolset²⁶. This approach implements a sliding window technique with overlap compensation to prevent artifacts at tile boundaries²⁷. TTP automatically sizes tiles (TTP_Tile_image_

size), splits the image into batches (TTP_Image_Tile_Batch) for parallel latent-space processing, and then reassembles them (TTP_Image_Assy), delivering a high-detail image while keeping memory usage in check. The tiling strategy ensures global consistency by maintaining overlapping regions and applying seamless blending algorithms during reconstruction²⁸.

Output and Post-Production. The 4K PNG sequence generated by the Flux pipeline is brought into After Effects at 24fps for grading, depth-of-field, and lighting (Lumetri Color, Camera Lens Blur). The PNG format ensures lossless quality preservation throughout the post-production pipeline²⁹. The graded master is output to DCI 4K (4096×2160 , 24 fps, Rec. 709 Gamma 2.4) following digital cinema standards³⁰.

Color grading utilizes the Rec. 709 color space, which provides standardized color reproduction for high-definition television and digital cinema applications³¹. The Lumetri Color panel implements industry-standard color correction algorithms including lift-gamma-gain adjustments and HSL (Hue, Saturation, Lightness) curves for precise color manipulation³². Camera Lens Blur applies depth-of-field effects using circle of confusion calculations to simulate authentic optical characteristics³³.

The final output is further polished in Topaz Video AI³⁴ using Chronos 2-4 \times frame interpolation. Chronos employs deep learning-based motion estimation and frame synthesis algorithms to generate intermediate frames, effectively increasing temporal resolution while maintaining motion coherence³⁵.

3.3 Presentation format

The piece is presented in two formats: (1) A 4K DCP version (24/50 fps, REC.709, 5.1 surround sound) designed for theatrical or festival screenings; (2) A gallery edition, looped from the same final master, projected in 4K with four-channel audio. Both formats maintain identical color grading and dynamic range, ensuring consistent presentation across black-box theaters and white-cube gallery settings.

4 Discussion

First, The potential of real-time image synthesis should be evaluated within a human-machine collaboration frame. As Mar Canet Sola notes, every “automatic” frame still requires the artist to architect the pipeline, select and tune models, and make iterative aesthetic and ethical calls [?]. AI is not a one-click oracle; it resembles developer fluid or pigment—only when the creator controls prompts, weights, and workflow does it speak a personal visual language. Just as photographers rely on framing, metering, and dark-room work,

²³Liu, X., et al. (2022). Flow matching for generative modeling. arXiv preprint arXiv:2210.02747.

²⁴Wang, X., et al. (2021). Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1905-1914).

²⁵Wang, X., et al. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) (pp. 63-79).

²⁶https://github.com/TTPPlanetPig/Comfyui_TTP_Toolset

²⁷Zhang, K., et al. (2019). Aim 2019 challenge on constrained super-resolution: Methods and results. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) (pp. 3565-3574).

²⁸Vaswani, A., et al. (2017). Attention is all you need. Advances in neural information processing systems, 30.

²⁹Boutell, T. (1997). PNG (Portable Network Graphics) specification version 1.0. RFC 2083.

³⁰Digital Cinema Initiative (2012). Digital Cinema System Specification Version 1.2.

³¹ITU-R Recommendation BT.709-6 (2015). Parameter values for the HDTV standards for production and international programme exchange.

³²Fairchild, M. D. (2013). Color appearance models. John Wiley & Sons.

³³Ray, S. F. (2002). Applied photographic optics: lenses and optical systems for photography, film, video, electronic and digital imaging. Focal Press.

³⁴<https://www.topazlabs.com/>

³⁵Niklaus, S., Mai, L., & Liu, F. (2017). Video frame interpolation via adaptive separable convolution. In Proceedings of the IEEE International Conference on Computer Vision (pp. 261-270).

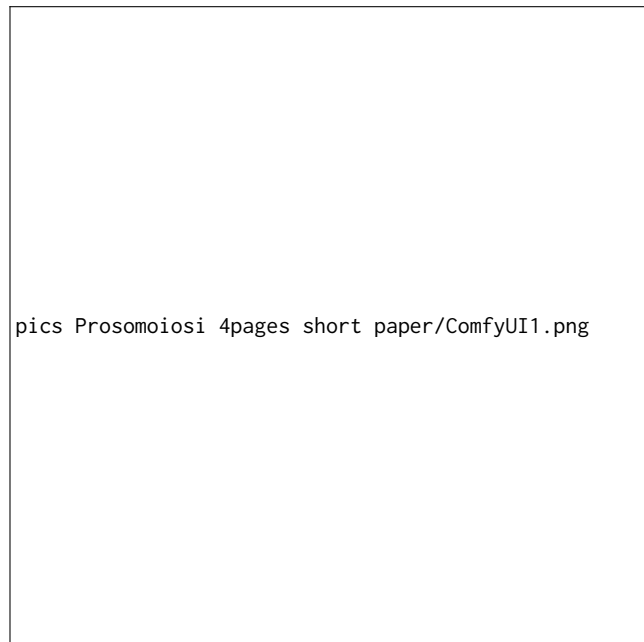


Figure 4: Overview of comfyUI upscaling workflow

real-time diffusion relies on human calibration of rhythm, light, and meaning. Therefore, before choosing live over offline generation, we must ask whether a project truly benefits from the transparency and improvisation of on-stage human-AI performance, rather than assuming the tool can replace authorship.

Elon Musk’s first-principles dictum—strip a problem to its basics, then build up [?]—shows what media art often gets wrong: choosing video or AI for fashion, then grafting on meaning. Start with the idea, then pick or invent the medium that enacts it; without that alignment, later polish is pointless.

Lightweight open-source toolkits such as Vadim Epstein’s SDFu³⁶ let artists shape real-time diffusion instead of merely triggering it. Epstein demonstrates this in live sets and in the short film *The Poem* [? ?], showing that “one-click” AI is a myth: prompt design, parameter tuning, and aesthetic judgment remain firmly human tasks. Real-time synthesis thus becomes less a technical shortcut than a new stage for co-authoring perception and authorship.

5 Conclusion

This paper has argued that contemporary generative AI practice requires more than technical novelty; it demands a principled reconciliation between medium and concept. Building on Musk’s call for first-principles reasoning, we reframed medium choice as a foundational decision rather than a cosmetic afterthought. The notion of medium alignment articulated here positions the operative logic of a medium—its algorithms, temporalities, and interfaces—as an integral layer of meaning production. Through the case study of *Prosomoiosi* (Simulation) we demonstrated how a TouchDesigner–StreamDiffusion pipeline, augmented by live parameter editing and Flux up-scaling, can turn real-time image synthesis

into both method and message. The work exposes the algorithmic contingencies of diffusion while inviting audiences to negotiate authorship in the very moment of visual emergence. Our comparative discussion showed that real-time and non-real-time pipelines offer distinct affordances—immediacy and performativity versus reflective plasticity—and that the critical task is to select the approach that most faithfully embodies a project’s conceptual agenda. Medium alignment thus emerges as a transferable design heuristic for artists and researchers navigating an era in which technical frameworks evolve faster than critical discourse. By treating first-principles reasoning and real-time generativity not as ends in themselves but as instruments of conceptual clarity, future practices can move beyond tool-centred spectacle toward works that disclose, rather than disguise, the politics of their own making.

Code availability and demo

The complete ComfyUI–TouchDesigner workflow used in this case study is openly available on GitHub, allowing readers to reproduce and extend our results: <https://github.com/UninstallAll/MAAPIIcursor> (see the workflow/ folder).

Acknowledgements

Acknowledgments have been removed for anonymous review and will be restored in the camera-ready version.

³⁶<https://github.com/eps696/SDFu>

Temporary page!

L^AT_EX was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because L^AT_EX now knows how many pages to expect for this document.