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

Shuai Liu
Academy of Media Arts Cologne (KHM)
Cologne, Germany
shuai.liu@khm.de

Mar Canet Sola
BFM, Tallinn University, Estonia
Academy of Media Art Cologne(KHM), Germany
mar.canet@gmail.com

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

Abstract

This paper examines the audiovisual artwork *Prosomoíosi (Simulation)*, advancing the concept 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 transparent live diffusion pipeline serves as both method and message, enabling audiences to observe how text prompts, stochastic noise, and performer input co-evolve on screen. By tracing precedents from early interactive installations to contemporary AI-driven works, the paper positions Prosomoíosi within a lineage that challenges tool-centric spectacle and renegotiates 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 explicitly reveal, rather than conceal, their underlying political dimensions.

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, generative artificial intelligence (AI) has rapidly transitioned from research laboratories to mainstream film production workflows. In 2024, the Academy of Motion Picture Arts and Sciences explicitly permitted films containing AI-generated footage to compete for Oscars—a symbolic milestone reflecting both Hollywood's recognition of this technological paradigm shift and a relaxation of traditional production standards. That same year saw the launch of the Runway AI Film Festival (New York and Los Angeles) and the RAIN Film Fest (Barcelona), signaling AI-based image creation's evolution from niche artistic practice to established ecosystem. The inaugural Asian AI Film Festival (HKUST AI Film Festival) in 2025 further confirmed this trend's global expansion and cultural diversification.

Technically, breakthroughs in generative models for scriptwriting, storyboarding, image synthesis, and long-form video generation have created unprecedented opportunities for real-time iteration. Text-to-video models like Runway Gen-2[1], OpenAI's Sora[2], and Google's Veo[3]—built on large-scale spatiotemporal diffusion frameworks—now produce controllable, high-definition output with narrative coherence. meanwhile, text-to-image models like Stable Diffusion[4], Flux[5], and Midjourney[6], 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 enhance production efficiency but also introduce a *machine vision* cultural grammar that co-evolves with audience perception.

Yet this technological progress reveals a critical contradiction: when artists adopt new media while retaining old conceptual frameworks—or vice versa—a misalignment emerges between medium and idea. The consequences typically manifest as:

(1) The new technology being reduced to a mere *accelerator*, unable to contribute meaningfully to the work's intent;

(2) Conceptual depth being sacrificed to the tool's default paradigm, resulting in demonstrative or techno-virtuosic superficiality;

(3) Audiences perceiving primarily a *technological spectacle* rather than the work's thematic core, thereby diluting its artistic statement.

Generative AI is rapidly reshaping film and art production, yet conceptual thinking has not kept pace, widening the gap between medium and idea. Through the case study of *Prosomoíosi (Simulation)*, we propose *medium alignment*: aligning a project's theme with its medium's operative logic from inception, making the algorithmic pipeline a visible narrative layer rather than hidden infrastructure. This proposal echoes Baudrillard's model-first simulation logic [?], Bolter and Grusin's remediation paradigm [?], and Manovich's analysis of algorithmic storytelling [?]. *Prosomoíosi* literalizes simulation: real-time images generated in TouchDesigner[7] via StreamDiffusion[8] (driven by keyframed parameters) are tiled and upscaled to 4K using ComfyUI and the 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 [9], 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 [10] *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 [11] 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" [?].

---

[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

[9] https://interface.ufg.ac.at/christa-laurent/

[10] https://www.memo.tv/

[11] http://var-mar.info

pics Prosomoiosi 4pages short paper/compare.jpg

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

This interrogation extends to works like Kevin Abosch's *Am I* documentary [?][12] and institutional critiques such as Nora Al-Badri's *The Post-Truth Museum* (2021–23) [?][13], which uses deepfakes to fabricate museum directors' speeches talking about decolonizing their museums. Recent works further explore AI's cultural implications: Paul Trillo's [14] Ars Electronica Golden Nica-winning *Washed Out "The Hardest Part"* (2024) demonstrates the potential of AI music videos with an impossible continuous traveling shot [? ?]. While Erin Robinson's [15] *XXX Machina* (2025) and Martyna

Marciniak's [16] *Anatomy of Non-Fact. Chapter 1: AI Hyperrealism* (2025) interrogate AI's role in constructing desire and documentary truth, respectively [? ?]. Together, these works—spanning interactive systems, institutional critique, and synthetic media—illustrate AI's evolution from celebratory tool to critical medium for examining how algorithms mediate.

3 Case Study : Prosomoiosi (Simulation)

3.1 Artwork Description

Prosomoíosi (Simulation) is a real-time generative video work that advances "medium alignment" to probe how technological regimes overwrite memory and identity. The work aims to (1) reveal how new media overprint cultural memory and (2) expose AI synthesis's algorithmic selectivity through a reflexive TouchDesigner–StreamDiffusion pipeline with real-time editing and ComfyUI upscaling. By allowing text prompts, noise, and performer input to co-evolve, the piece shifts image production from depiction to ontological negotiation.

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: This "from code to camera" trajectory finds expression in Martyna Marciniak's *Anatomy of Non-Fact, Chapter 1: AI Hyperrealism* (2025), which follows Baudrillard's "proving the real through the imaginary" by generating synthetic imagery that questions fact-fiction boundaries [?][17], and Erin Robinson's *XXX Machina* (2025), which uses algorithmic processes to create continuously evolving outputs that embrace computational unpredictability [?][18].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. 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 to TouchDesigner for real-time generation.

Real-Time Diffusion Generation (TouchDesigner + StreamDiffusion). Hardware platform: AMD Ryzen 9 7950X / NVIDIA RTX 4090 (24 GB VRAM). In TouchDesigner, the StreamDiffusion node developed by dotsimulate[19] is loaded. After configuring the project's dependencies, every video frame undergoes image-to-image (img2img) diffusion. Models such as Stable Diffusion XL, Stable Diffusion 1.5, and their LoRAs are switched dynamically to achieve desired artistic effects. StreamDiffusion parameters—Prompt, CFG Scale, 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. The workflow also integrates ControlNet's HED (Holistically Nested Edge Detection) branch within TouchDesigner; by parameterizing the edge-strength weight, the diffusion process is further constrained to converge on scene geometry and spatial alignment.

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× NMKD-Siax super-resolution pass; this tile-based routine balances memory, accommodates large canvases, and boosts texture and edge detail for later stages.

To reduce VRAM load at high resolutions, we use the Tile-to-Patch (TTP) toolset[20]. 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.

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 graded master is output to DCI 4K (4096 × 2160, 24 fps, Rec. 709 Gamma 2.4) and polished in Topaz Video [21] AI—Chronos 2-4× frame-rate.

---

[16]https://www.martyna.digital/
[17]https://calls.ars.electronica.art/2025/prix/winners/19504/
[18]https://erinrobinson.xyz/artworks/xxx-machina

[19]https://dotsimulate.com/
[20]https://github.com/TTPlanetPig/Comfyui_TTP_Toolset
[21]https://www.topazlabs.com/
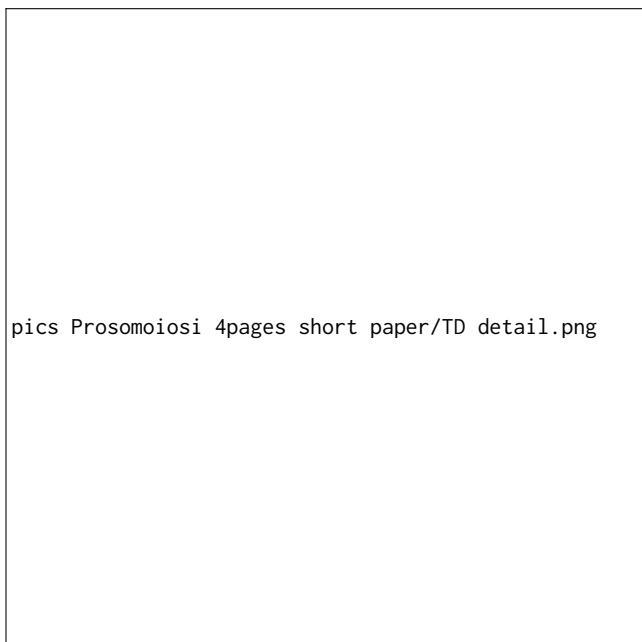
pics Prosomoiosi 4pages short paper/TD detail.png

Figure 3: Screen capture of the TouchDesigner workflow: red – Animation COMP–related modules; green – StreamDiffusion node; yellow – ControlNet
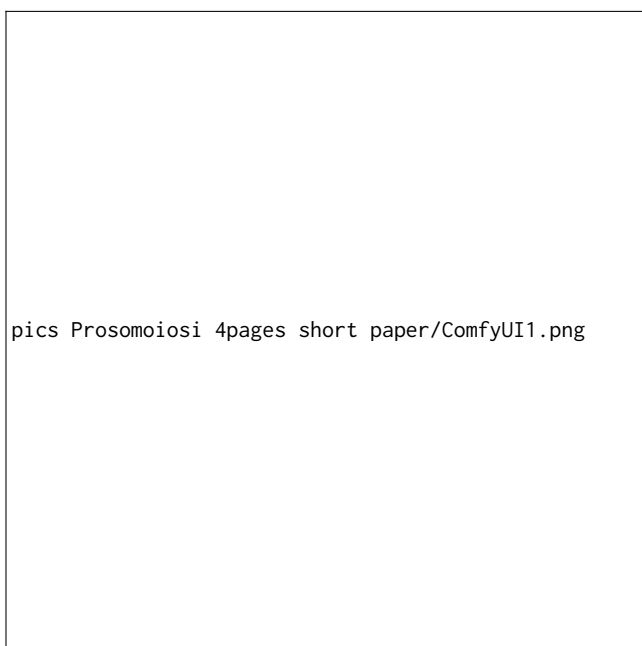
pics Prosomoiosi 4pages short paper/ComfyUI1.png

Figure 4: Overview of comfyUI upscaling workflow

3.3 Presentation format

The work is presented as: (1) 4K DCP for theatrical screenings (24/50 fps, 5.1 surround); (2) looped 4K gallery projection with quadraphonic audio. Both maintain identical grading for consistent black-box/white-cube presentation.

4 Discussion

First, real-time image synthesis should be evaluated within a human–machine collaboration frame. As Guljajeva et al. note, every "automatic" frame still requires the artist to architect the pipeline, select models, and make aesthetic decisions [?]. AI resembles developer fluid or pigment—only when creators control prompts, weights, and workflow does it speak a personal visual language. Just as photographers rely on framing and darkroom work, real-time diffusion requires human calibration of rhythm, light, and meaning. Therefore, we must ask whether a project truly benefits from the transparency of on-stage human-AI performance, rather than assuming tools 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 SDfu[22] 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 first-principles reasoning, we reframed medium choice as foundational rather than cosmetic. The notion of medium alignment positions the operative logic of a medium—its algorithms, temporalities, and interfaces—as an integral layer of meaning production. Through the case study of Prosomoíosi (Simulation) we demonstrated how a TouchDesigner–StreamDiffusion pipeline can turn real-time image synthesis into both method and message. The work exposes algorithmic contingencies while inviting audiences to negotiate authorship during visual emergence. Our discussion showed that real-time and non-real-time pipelines offer distinct affordances, 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 navigating an era where technical frameworks evolve faster than critical discourse. By treating real-time generativity as an instrument 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

---

[22]https://github.com/eps696/SDfu

and extend our results: https://github.com/UninstallAll/MAAPIIcursor (see the `workflow/` folder).

Acknowledgements