From Code to Camera:

The Making and Meaning of Prosomoiosi (Simulation),

an AI Documentary Film

Authors:

Shuai Liu

Academy of Media Arts Cologne (KHM), Cologne, Germany

Email: shuai.liu@khm.de

Mar Canet Sola

BFM, Tallinn University, Estonia

Academy of Media Arts Cologne (KHM), Germany

Email: mar.canet@gmail.com

ABSTRACT

This paper examines the audiovisual artwork Prosomoiosi (Simulation), advancing the concept of "medium alignment" to critique how successive simulation technologies overwrite cultural memory. Grounded in media-archaeological theory, the study argues that concepts must be articulated through media whose operative logic remains visible, 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 Prosomoiosi 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.

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

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, OpenAI's Sora, and Google's Veo—built on large-scale spatiotemporal diffusion frameworks—now produce controllable, high-definition output with narrative coherence. Meanwhile, text-to-image models (Stable Diffusion, Flux, Midjourney) and large language models (LLMs) for script development are reshaping traditional pre-production workflows. 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.

As generative AI transforms film and art production, conceptual thinking has failed to keep pace, widening the medium-idea gap. Through the case study of Prosomoiosi (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 approach resonates with Baudrillard's simulation theory, Bolter and Grusin's remediation paradigm, and Manovich's algorithmic storytelling analysis.

Prosomoiosi literalizes simulation: real-time images generated in TouchDesigner via StreamDiffusion (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 witness algorithms rewriting the image itself. We argue such medium alignment enables generative video to transcend tool fetishism, creating a dynamic arena for renegotiating memory, history, and subjectivity.

REFERENCE ARTWORKS

To contextualize our research, we examine artworks tracing the evolution from early interactive systems to contemporary AI-driven perception critiques. A foundational example is A-Volve (1994) by Christa Sommerer and Laurent Mignonneau, where visitors' drawings become virtual creatures, foregrounding artificial life and embodied interaction. Decades later, deep learning has reconfigured this focus—artists now use AI to interrogate seeing itself. Memo Akten's Learning to See (2017) employs neural networks to reveal how models reconstruct reality through biased training data filters, explicitly questioning perception-truth relationships. Scott Eaton's Entangled II (2019) processes abstract footage into morphing quasi-human figures using a bespoke neural network, examining machine perception aesthetics and human pareidolia. Taking data critique further, POSTcard Landscapes from Lanzarote (2022) by Varvara & Mar uses StyleGAN2 trained on distinct datasets to generate opposing visual realities of one location—tourist versus local perspectives—commenting on how curated data shapes cultural identity and memory (echoing Urry's "tourist gaze"). The authors extended this trajectory in a 2024 work exploring AI-human symbiosis through real-time generation (see Figure 2 for quality comparisons). Together, these works illustrate a paradigm shift from celebrating artificial life interaction to critically examining how AI models mediate reality construction.

CASE STUDY: PROSOMOIOSI (SIMULATION)

ARTWORK DESCRIPTION

Prosomoiosi (Simulation) is a real-time generative video work that, through a media-archaeological lens, employs "medium alignment" to probe how technological regimes overwrite memory and identity. The work pursues two interconnected aims:

1. To reveal, through stratified simulation logics, how new media silently overwrite cultural remembrance.

2. To expose AI image synthesis' algorithmic selectivity and symbolic power via a reflexive TouchDesigner-StreamDiffusion pipeline with real-time multiprompt editing and ComfyUI upscaling (flux-dev f16 model).

By enabling text prompts, stochastic noise, and live performer input to co-evolve in exhibition space, the piece transforms image production from depiction to ontological negotiation, inviting audiences to reconsider human-machine creativity dynamics.

Simulation here constitutes not mere copying but a power-laden rewriting through technology-archives-algorithms interplay. Wolfgang Ernst observes archival temporality shapes memory while editing the past; Benjamin anticipated aura's erosion through mechanical reproduction; Baudrillard warned of simulacra replacing reality. Tracing a "clay to code" trajectory—from sand tables and armillary spheres to deep learning and GANs—modeling technologies have evolved from cognitive tools to truth arbiters: networks like GPT-4 and ESM-2 autonomously generate worlds from latent space and even guide human decisions, reshaping human-reality relations. Video games now dominate as simulation media, their "fourth wall" breaches placing players in hybrid spaces where procedural rhetoric infuses interaction with symbolic capital and ideology—extending McLuhan's "the medium is the message" and Bourdieu's symbolic power analysis. Educational/political simulations (SimCity, PeaceMaker) and Lorenz-style chaotic systems further demonstrate simulation's profound behavioral and cognitive influence. Confronting algorithmic culture's media mutations requires medium alignment—expanding AI alignment into perceptual-symbolic calibration at the medium level, positioned between Stiegler's "third memory" and Yuk Hui's "cosmic technics" to foster human-machine-media symbiosis.

PRODUCTION WORKFLOW

The real-time generation pipeline comprises four stages:

1. Material Preparation

Source footage (public-domain vintage films and live gameplay recordings) is transcoded to 24 fps H.264 in Adobe Premiere Pro. For optimized RTX 4090 diffusion processing, footage is slowed to 4-14 fps (0.2x-0.5x speed) and streamed via NDI to TouchDesigner.

2. Real-Time Diffusion Generation (TouchDesigner + StreamDiffusion)

Hardware: AMD Ryzen 9 7950X/NVIDIA RTX 4090 (24GB VRAM)

The StreamDiffusion node (dotsimulate) performs image-to-image diffusion using models like Stable Diffusion XL/1.5 and their LoRAs, dynamically switched for artistic effects

Parameters (Prompt, CFG Scale, Steps, Denoise) are keyframe-controlled via TouchDesigner's Animation COMP for narrative rhythm

ControlNet's HED edge detection constrains diffusion convergence on scene geometry

3. High-Resolution Upscaling (ComfyUI Flux Workflow)

Images resized to 1024×1024 for VRAM-optimized processing

4x NMKD-Siax super-resolution applied

Tile-to-Patch (TTP) toolset enables memory-efficient high-detail output:

TTP\_Tile\_image\_size: automatic tile sizing

TTP\_Image\_Tile\_Batch: parallel latent-space processing

TTP\_Image\_Assy: seamless reassembly

4. Post-Production

4K PNG sequences imported to After Effects (24fps)

Grading, depth-of-field, lighting (Lumetri Color, Camera Lens Blur)

Final output: DCI 4K (4096×2160, 24 fps, Rec.709 Gamma 2.4)

Optional Topaz Video AI Chronos 2-4x frame interpolation

PRESENTATION FORMAT

Two display formats ensure versatile exhibition:

1. Theatrical: 4K DCP (24/50 fps, REC.709, 5.1 surround) for festival screenings

2. Gallery: Looped 4K projection with quadraphonic audio

Both maintain identical color grading and dynamic range for consistent black-box/white-cube presentation.

DISCUSSION

Real-time image synthesis should be evaluated through human-machine collaboration frameworks. As Mar Canet Sola notes, each "automatic" frame still requires pipeline architecture, model selection/tuning, and iterative aesthetic/ethical decisions. AI isn't a one-click oracle—it becomes expressive only when creators control prompts, weights, and workflows. Just as photographers master framing, metering, and darkroom techniques, real-time diffusion demands human calibration of rhythm, light, and meaning. Thus, choosing live generation requires determining whether a project genuinely benefits from on-stage human-AI performance transparency and improvisation—not assuming tools can replace authorship.

Elon Musk's first-principles approach—deconstructing problems to fundamentals before rebuilding—highlights what media art often gets wrong: selecting video or AI for fashion, then grafting meaning. Instead, start with ideas, then choose (or invent) media that enact them. Without this alignment, post-hoc polish proves futile.

Lightweight open-source toolkits like Vadim Epstein's SDfu empower artists to shape rather than merely trigger real-time diffusion. Epstein's live sets and short film The Poem demonstrate that "one-click" AI is mythical—prompt design, parameter tuning, and aesthetic judgment remain firmly human. Real-time synthesis thus becomes less a technical shortcut than a new stage for co-authoring perception and meaning.

CONCLUSION

This paper has argued that contemporary generative AI practice demands principled medium-concept reconciliation beyond technical novelty. Building on Musk's first-principles reasoning, we reframed medium choice as foundational rather than cosmetic. The proposed medium alignment positions a medium's operative logic—its algorithms, temporalities, and interfaces—as integral to meaning production.

Through Prosomoiosi (Simulation), we demonstrated how a TouchDesigner-StreamDiffusion pipeline with live parameter editing and Flux upscaling can transform real-time synthesis into both method and message. The work exposes diffusion's algorithmic contingencies while inviting authorship negotiation during visual emergence. Our discussion showed real-time and offline pipelines offer distinct affordances—immediacy/perform