

Fabian Offert<sup>1</sup>

## Five Theses on the End of AI Art

1. This short text makes a technical, media-theoretical, and likely also controversial, statement about the future of artificial intelligence as an artistic medium. The statement reads: with the switch from generative adversarial networks (GANs) to transformer/diffusion-based models (TDMs), AI art<sup>2</sup> has come to an end. Given that more people than ever use artificial intelligence systems to produce texts and images, the statement appears counter-intuitive. We thus need to point out that the statement is historical in nature. The argument is not that there is no more art made with the help of artificial intelligence but that **the switch from GANs to TDMs has aesthetic implications that are so significant that it becomes impossible to understand AI art pre and post 2021<sup>3</sup> as the same genre**. Media famously determine our situation. In the case of AI art, this determination is all-encompassing.

2. I have argued elsewhere<sup>4</sup> that AI art is essentially sculpture: a subtractive, not additive artistic process. AI art images are not ‘generated’, they are ‘discovered’ by navigating the vast latent space provided by both GANs and TDMs. ‘Making’ an AI art image, in that sense, means throwing away all others. When Michelangelo states that all that is necessary to create a sculpture is to chisel away the superfluous material from a block of marble, he describes a process of guided discovery in which artistic intuition is realized materially. For a latent space to become an artistic medium, there has to exist some principle of navigation that facilitates this exact kind of guided discovery. This does not necessarily imply the absence of randomness – otherwise we would not speak of “discovery”. But it does imply the existence of some structure, even if it is one with infinite junctions, of some intention, even if it is the intention to reach maximum arbitrariness. In both GAN-based and TDM-based AI art, **the principle of navigation is where artistic experimentation takes place<sup>5</sup>**.

3. Crucially, this principle of navigation has to take into account that **the discovery of images in latent space is actually a process of translation**. While, on the one hand, the notion of “generative model” makes clear that nothing is ever already there, the popular assumption is that latent spaces are “filled” with images. This is false. Rather, latent spaces are, in the first instance, just a region of high-dimensional space, arbitrarily sectioned off according to model architecture and hyperparameter selection. An image comes into existence if and only if a point in this space, represented by a vector, is processed by the model during the inference stage. This means that the discovery of images in latent space must proceed iteratively. Navigation, in large visual models, is an optimization problem. A point is selected, an image is generated, another point is selected, another image is generated, and so on.

---

<sup>1</sup> Fabian Offert is Assistant Professor for the History and Theory of the Digital Humanities at the University of California, Santa Barbara, and principal investigator of the “AI Forensics” project (2022-2025), funded by the Volkswagen Foundation.

<sup>2</sup> This text is exclusively concerned with art created with, not about, artificial intelligence.

<sup>3</sup> The release of CLIP arguably is the technical backbone of the aesthetic paradigm change discussed in this text.

<sup>4</sup> Offert, Fabian. “KI-Kunst als Skulptur”. In: *KI-Realitäten. Modelle, Praktiken und Topologien maschinellen Lernens*. Ed. by Richard Groß and Rita Jordan. Transcript, 2023 (forthcoming).

<sup>5</sup> Obviously, the principle of navigation is not the only place of artistic experimentation. Both in early GAN-based art and in current experiments with Stable Diffusion, the training or, more often, fine-tuning of models is an aesthetically relevant activity, as artists like Helena Sarin demonstrate. Moreover, a generated image, a digital file, by itself, does not make a generative work – we are intentionally disregarding this more traditional issue of AI art in this short text.

4. **In GANs, the principle of navigation is geometric.** Given that GANs are unimodal models – they know images and images only – there exists no interface that would facilitate the discovery of images other than latent space itself. Importantly, the mathematics of latent space are non-Euclidean. As suggested by the term “curse of dimensionality”, in high dimensional spaces, intuition breaks down. For GANs, this means that the discoverability of latent space is significantly constrained. Artists had to come up with ways to address or circumvent these constraints<sup>6</sup>, and most artistic ‘innovations’ before 2021 consist of such geometric principles of navigation, including both complex explorations of latent space topology, as well as simple ‘hacks’ to achieve effects like ‘acceleration’ and ‘deceleration’ in latent space interpolation videos.

5. **In TDMs, the principle of navigation is semantic.** TDMs (and CLIP-guided GANs) are multimodal models. They offer a natural language ‘interface’ based on CLIP<sup>7</sup> that allows the user to address, and thus navigate, latent space symbolically. Consider signal processing as an analogy. GANs are like electronic music, where meaning is created by painstakingly layering sinusoidal frequencies by hand. Meaning only emerges once a certain complexity is reached. In TDMs, a symbolic system that already encodes meaning – natural language – is used to guide an instrument that does the hard material work for us. TDMs, then, are like classical music, where the act of creativity lies in the composition<sup>8</sup>. Importantly, none is inherently “better” than the other – but both are different enough to warrant their classification as two entirely different activities. **The end of AI art (as signal processing), in other words, marks the beginning of AI art (as composition).**

---

<sup>6</sup> One early method of ‘evaluation’ for GANs was ‘finding’ Leonardo DaVinci’s *Mona Lisa* in the GAN latent space. The closer to the actual painting the generated image appeared, the better the respective principle of navigation was assumed to work.

<sup>7</sup> Radford, Alec, Jong Wook Kim, et al. “Learning Transferable Visual Models from Natural Language Supervision.” In International Conference on Machine Learning (ICML), 8748–63, 2021.

<sup>8</sup> This is of course an oversimplification for the sake of analogy and not intended to challenge the existence of electronic instruments on the one hand, or the aesthetic relevance of musical interpretation on the other.