Lev Manovich

The AI Brain in the Cultural Archive

Notes on Refik Anondol Unsupervised AI art project for MoMA, and connections between generative AI and artistic methods of the 19th and 20th century

(Shorter version of this text published in MoMA magazine, July 21, 2023. Another version was included in Chapter 5 of "Artificial Aesthetics" book.)

We now know a lot more about human creativity and how it works than we do about AI creativity. Since the 1950s, many theories of creativity have been established in philosophy, psychology, cognitive science, and other domains, and multiple types of creativity have been described. We are likely to do the same with AI creati1vity over time, but we are not there yet. After training on trillions of text pages or billions of images and their descriptions from the web, artificial neural networks can generate fresh texts and visuals on the level of highly competent professional writers, artists, photographers, or illustrators. These capacities of these systems s are distributed over trillions of connections between billions of artificial neurons rather than determined by standard algorithms. In other words, we developed a technology that, in terms of complexity, is extremely similar to the human brain. We don't fully grasp how our AI technology works, just as we don't fully comprehend human intellect and creativity.

The current generation of Generative AI tools, such as GPT and Stable Diffusion, have been trained on very large and diverse datasets consisting from billions or even trillions of individual texts, or image and text pairs. It is, however, equally interesting to limit the training data set to a specific area within the larger space of human cultural history, or to a specific set of artists from a specific historical period. *Unsupervised* by Refik Anadol Studio (2022) is an AI art project that exemplifies these possibilities

The project uses a neural network trained on the image dataset of tens of thousands of artworks from the MoMA collection. This collection, in my opinion, is one of the best representations of the most creative and experimental period in human visual history—the hundred years of modern art (1870–1970)—as well as many important examples of artistic explorations in the subsequent decades. It captures modernist artists' feverish and relentless experiments to create new visual and communication languages and "make it new."

Generative AI, Modernism and Tradition

On the surface, the logic of modernism appears to be diametrically opposed to the process of training generative AI systems. Modern artists desired to depart from classical art and its defining characteristics, such as visual symmetry, hierarchical compositions, and narrative content. In other words, their art was founded on a fundamental rejection of everything that had come before it (at least in theory, as expressed in their manifestos). Neural networks are trained in the opposite manner, by learning from historical culture and art created up to now. A neural network is analogous to a very conservative artist studying in the "meta" "museum without walls" that houses historical art.

But we all know that art theory and art practice are not the same thing. Modern artists did not completely reject the past and everything that came before them. Instead, modern art developed by reinterpreting and copying images and forms from old art traditions, such as Japanese prints (Vincent van Gogh), African sculpture (Picasso), and Russian icons (Malevich).

Thus, the artists only rejected the dominant high art paradigms of the time, realistic and salon art, but not the rest of human art history. In other words, modernism was deeply historicist; rather than inventing everything from scratch, it innovated by adapting certain older aesthetics to contemporary art contexts.

(In the case of geometric abstract art that emerged in the early 1910s, similar images were already widely used in experimental psychology to study human visual sensation and perception. For a detailed analysis of these relations, see Paul Vitz and Arnold Glimcher, Modern art and Modern Science: The Parallel Analysis of Vision, 1983.)

When it comes to artistic AI, we should not be blinded by how these systems are trained. Yes, artificial neural networks are trained on previously created human artifacts. However, their newly generated outputs are not mechanical replicas or simulations of what has already been created. In my opinion, these are frequently genuinely new cultural artifacts with previously unseen content, aesthetics, or styles.

Of course, simply being novel does not automatically make something culturally or socially interesting or significant. Indeed, many definitions of "creativity" agree on this point: it is the creation of something that is both original and worthwhile or useful.

However, estimating what percentage of all novel artifacts produced by generative AI are also "useful" (or "meaningful") for a larger culture is not a feasible project at this time. For one thing, I am not aware of any systematic effort to use such systems to "fill in," so to speak, a massive matrix of all content and aesthetic possibilities by providing millions of specifically designed prompts. Instead, it is likely that, as in every other area of popular

culture, only a small number of possibilities are realized over and over by millions of users, leaving a long tail of other possibilities unrealized. So, if only a tiny fraction of the vast universe of potential AI creations is being realized in practice, we can't make broad statements about the originality or utility of the rest of the universe.

Media Accumulations: AI and Database Art

Some AI artists, such as Anna Ridler, Sarah Meyohas and Refik Anadol utilized in their works neural networks trained on specific datasets. Other artists, designers, architects, and many other users of AI tools use already pre-trained networks released by other companies, organizations, or research institutions, and then fine tune them on their own datasets.

For example, young artist Lev Pereulkov fine tuned Stable Diffusion model 2.1 on his dataset of 40 paintings by famous Russian conceptual artists (Erik Bulatov, Ilya Kobakov, Irina Nakhova and others). The series of images entitled «искусственные эксперименты 1–10» (artificial experiments 1–10) generated by Pereulkov using this customized net reflects not only stylistic features of these artists but also their particular surreal and absurd semantics.

Today, the majority of ordinary people and creative professionals who use generative media tools use them as is, without further fine tuning them on their own large image collections. This may change in the future as the training networks using our own dataset become easier and faster. (Some AI tools, such as Runway, allow you to "train" your own model using only 15–30 images.) But regardless, all new cultural artifacts made with generative AI follow one common principle.

These artifacts are not created from scratch like traditional drawings, sculptures, or paintings. They are also not the result of capturing sense phenomena such as photos, videos, or sound recordings. Instead, they are constructed from a large archive of other media artifacts.

This principle aligns generative AI with many earlier art genres and art making techniques.

For example, we can relate generative AI to film editing that emerged around 1898—and also to experimental collage films such as Bruce Conner's *A Movie* (1958) — or even earlier composite photography very popular in the 19th century.

Seeing projects such as *Unsupervised* or *artificial experiments 1–10* in the larger context of this very general media creation method and its historical variations will help us to more clearly see this and many other AI artworks as art objects engaged in dialogs with art of the past, rather than as purely technological novelties or works of entertainment.

Scanning the history of art, visual culture, and media for other prominent uses of this procedure, I see many relevant moments and periods. They are related to the current moment not only because artists working in these times used the procedure itself, but also because the reason for this use was the same in all cases. Throughout the 20th century, new types of large accumulations of cultural artifacts led artists to create new forms of art driven by these accumulations. Let me point out a few of these cases.

In the late 1990s and early 2000s, net artists created a number of works that reacted to the rapidly growing accumulation of cultural objects on the world wide web. For example, Health Bunting's _readme.html _(1999) is a web page containing the text of an article about the artist, with every word linked to an existing web domain corresponding to such a word. The Shredder by Mark Napier (1998) presents a dynamic montage of elements that make up numerous websites: images, texts, HTML code, and links.

Going back a bit earlier in time, we find a broad cultural paradigm that was also a response to the accumulation of historical art and culture. This paradigm is so-called post-modernism in art and design. Rejecting modernism's focus on novelty and breaking with the past, post-modern artists and designers celebrate bricolage, creating works consisting of quotations and references to art of the past. There are many possible reasons for the emergence of the post-modern paradigm in the 1960s-1980s and one of the proposed explanations is directly relevant here. The accumulation of earlier art and media artifacts in structured collections such as slide libraries, film archives, art history textbooks with lots of photos of the artworks, and other formats—where different historical periods, movements, and creators were positioned together — motivated postmodern artists to start creating collages from such references.

And what about earlier 20th century modernism ? While its overall emphasis was originality and newness, one of the procedures it developed in this search for novelty paradoxically was direct quotations from the vast universe of visual media that was rapidly expanding in the 1900–1920s. (For example, large headings and the inclusion of photos and maps made newspapers more visually impactful; new visually oriented magazines such as Times and Vogue were also started in this period.)

Responding to this visual turn in mass culture, in the early 1910s Georges Braque and Pablo Picasso started to incorporate actual fragments of newspapers, posters, wallpapers, or pieces of colorful fabrics in their paintings. A few years later, artists such as John Heartfield, George Grosz, Hannah Höch, Aleksander Rodchenko, and others developed another method of using mass media images: photo-collage.

Contemporary artworks that use neural networks trained on massive cultural databases such as Refik Anadol Studio's *Unsupervised* stand in the long tradition of making new art from accumulations of images and other media. In this way, these artworks made with

generative AI continue to expand the possibilities of art and its methods, especially the methods of what I called database art (see my article Database as a Symbolic Form, 1998.) This expansion includes the introduction of new methods for "reading" cultural databases. The double brain of Unsupervised's creators (i.e., the human brains of the studio members and the "artificial brain" of the neural net) does not quote from historical art or create media collages as earlier artists. Instead, it extracts patterns and relationships from tens of thousands of artworks collected by MoMA. It then uses everything it learned to generate the endlessly changing visual surface that interpolates between these artworks and their patterns in unpredictable ways. The surface oscillates between 2D and 3D, between pure abstraction and recognizable fragments of particular artworks, between gentle movements and dramatic visual explosions.

Looking at this work, I am reminded of the alien planet in Tarkovsky's *Solaris* (1972). Solaris synthesizes human-like "guests" from the memories of human inhabitants of the space station orbiting the planet. The surface of the planet also changes throughout the film, sometimes showing concrete images and then starting to form islands at the end. Somewhat similar to *Solaris*, *Unsupervise*d generates something that is both very familiar and also alien from all the thoughts of the artists who made the artworks that MoMA collected. Perhaps it is an early glimpse of the new forms of consciousness that will emerge between humans and the computer systems they engineered—and a future language of culture, which for now we are just learning how to read.