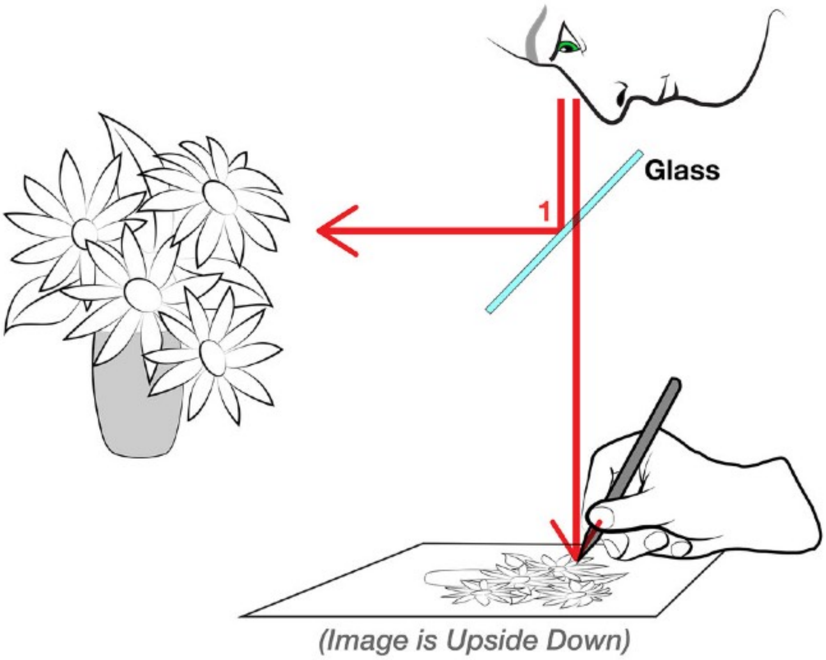
Lev Manovich

Seven Arguments about AI Images and Generative Media

Chapter 5 of *Artificial Aesthetics* by Lev Manovich and Emanuele Arielli. Published April 20, 2023.[1] 6,385 words. All book chapters: http://manovich.net/index.php/projects/artificial-aesthetics-book



Basic principle of camera lucida, optical drawing device widely used by artists and art students in the 19th century. Source: https://neolucida.com/history.

[1] Earlier version of one parts of this chapter will appear in *Diffusions – Taxonomy of Synthetic Imaginations in Architecture*, ed Matias del Campo, Willey, 2023; and an early shorter version of another part will be published in MoMA magazine, 2023. This chapter is written by me without use of any AI writing tools. In editing stage, I use quillbot tool that offer paraphrased versions of the selected sentences.

We appear to be in the the beginning of a true revolution in media creation: the rise of generative media. I've been using computer tools for art and design since 1984, and I've seen a few major media revolutions, including the introduction of Mac computers and desktop applications for media creation and editing, the development of photorealistic 3D computer graphics and animation, the rise of the web after 1993, and the rise of social media networks after 2006. The new AI generative media revolution appears to be as significant as any of them. Indeed, it is possible that it is as significant as the invention of photography in the nineteenth century or the adoption of linear perspective in western art in the sixteenth.

(If you are new to this topic, here is very brief history. Generative media revolution was in development for over 20 years. The first AI papers proposing that the vast unstructured web universe of texts, images and other cultural artifacts can be used to train computers to do various tasks appeared already in 1999-2001. In 2015 Google “deep dream” and “style transfer” methods attracted lots of attention: suddenly computers could create new artistic images mimicking styles of many famous artists. The release of DALL-E in January 2021 was another milestone: now computers could synthesize images from text description. Midjourney, Stable Diffusion, and DALL-E 2 all contributed to the acceleration of this evolution in 2022. Now synthetic images could have many aesthetics that ranges from photo-realism to any kind of physical or digital medium, including mosaics, oil paintings, street photography, or 3D CG rendering. The code for producing such images referred to as a "model" in the field of artificial intelligence was made public in August 2022, sparking a flurry of experiments and accelerating development.)

In this and the next chapters I will describe a number of characteristics of *visual generative media* in its current forms that I believe are particularly significant or novel. Some of my arguments also apply to generative media in general, but most focus on *visual media* - reflecting my own experience of using a few popular AI image tools such as Midjourney and Stable Diffusion (and sometimes also Runway ML) almost every day from middle of 2022 to early 2023. But first, let's define the main terms.

The Terms

In this text, "artist" or "creator" refers to any skilled person who creates cultural objects in any media or their combinations. The terms “generative media,” “AI media,” “generative AI,” and “synthetic media” are all interchangeable. They refer to the process of creating new media objects with deep neural networks, such as images, animation, video, text, music, 3D models and scenes, and other types of media. Neural networks are also used to generate specific elements and types of content, such as photorealistic human faces and human poses and movements, in addition to such objects. They can also be used in media editing, such as replacing a portion of an image or video with another content that fits spatially.

These networks are trained on vast collections of media objects already in existence. Popular artificial neural network types for media generation include diffusion models, text-to-image models, generative adversarial networks (GAN), and transformers. For the generation of still and moving images using neural networks, the terms image generation, synthetic image, AI image, and AI visuals can be used interchangeably.

Note that the word “generative” can be also used in different ways to refer to making cultural artifacts using any algorithmic processes (as opposed to only neural networks) or even a rule-based process that does not use computers. This is how the phrases “generative art” and “generative design” are typically used today in cultural discourses and popular media. In this chapter I am using “generative” in more restrictive way to designate deep network methods and apps for media generation that use these methods.

Note that the word "generative" can also be used in different ways to mean making cultural artifacts using any algorithmic process (not just neural networks) or even a rule-based process that doesn't use computers. This is how the terms "generative art" and "generative design" are often used in popular culture and the media today. I use "generative" in a narrower sense to refer to deep network methods to make media artifacts and apps that use these methods.

1. ’AI’ as a Cultural Perception

There is not one specific technology or a single research project called ‘AI’. However, we can follow how our cultural perception of this concept evolved over time and what it was referring to in each period. In the last fifty years, when an allegedly uniquely human ability or skill is being automated by means of computer technology, we refer to it as ‘AI’. Yet, as soon as this automation is seamlessly and fully successful, we tend to stop referring to it as an ‘AI case’. In other words, ‘AI’ refers to technologies and methodologies that automate human cognitive abilities and are starting to function but aren't quite there yet. 'AI' was already present in the earliest computer media tools. The first interactive drawing and design system, Ivan Sutherland's *Sketchpad* (1961-1962), had a feature that would automatically finish any rectangles or circles you started drawing. In other words, it knew what you were trying to make. In the very broad understanding just given, this was undoubtedly 'AI' already.

My first experience with a desktop paint program running on an Apple II was in 1984, and it was truly amazing to move your mouse and see simulated paint brushstrokes appear on the screen. However, today we no longer consider this to be 'AI'. Another example would be the Photoshop function that automatically selects an outline of an object. This function was added many years ago – this, too, is 'AI' in the broad sense, yet nobody would refer to it as such today. The history of digital media systems and tools is full of such 'AI moments' – amazing at first, then taken for granted and forgotten as 'AI' after a while. (In AI history books, this phenomenon is referred to as the 'AI effect'.) At the moment, 'creative AI' refers only to recently developed methods where computers transform some inputs into new media outputs (e.g., text-to-image models) and specific techniques (e.g., certain types of deep neural networks). However, we must remember that these methods are neither the first nor the last in the long history and future of simulating human art abilities or assisting humans in media creation.

2. "Make it New": AI and Modernism

After training on trillions of text pages or billions of art and photographic pictures taken from the web, neural networks can generate fresh texts and visuals on the level of highly competent professional writers, artists, photographers, or illustrators. These capacities of the AI systems nets 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 systems, 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 a AI art project that exemplifies these possibilities. The project uses neural networks 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 - 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."

Unsupervised, Refik Anadol Studio (2022). Selected frames from the animation.

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 (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, it 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 created in 1910s, these artists used images that were already widely used in experimental psychology to study human visual sensation and perception. For the detailed analysis of these relations between modern art and experimental psychology, 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 art and culture 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*.

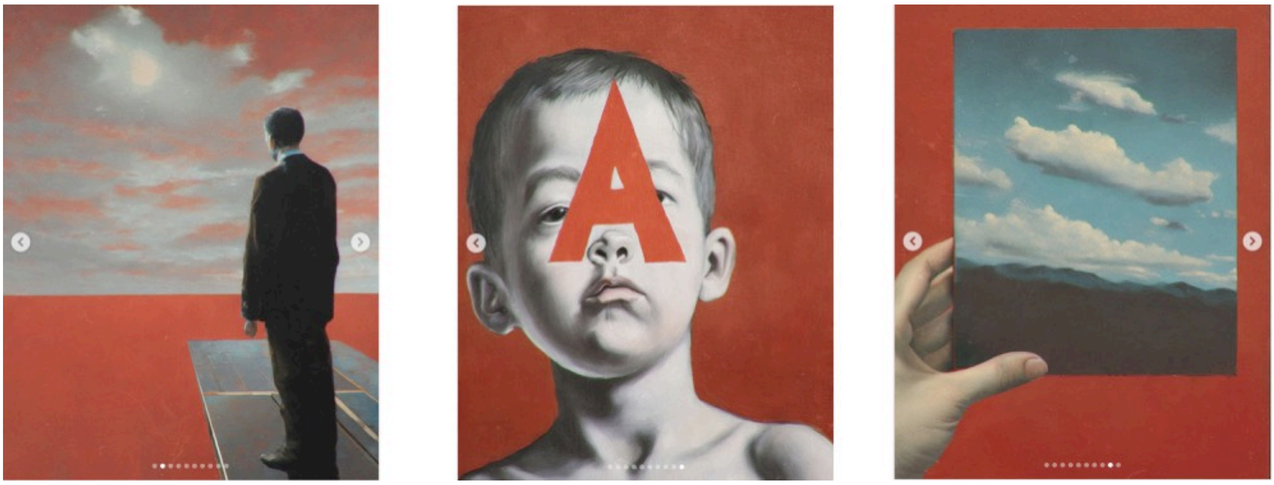
Of course, simply being novel does not automatically make something culturally or socially interesting or significant. Indeed, many definitions of "creativity" agree on this point: it is the creation of something that is both original and worthwhile or useful.

However, estimating what percentage of all novel artifacts produced by generative AI are also useful and/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 artifacts is being realized in practice, we can't make broad statements about the originality or utility of the rest of the universe.

3. Generative Media and Database Art

Some AI artists such Anna Ridler, Sarah Meyohas and Refik Anadol utilized in their works nets trained on specific datasets. Many other artists, designers, architects, and technologists use networks released by other companies or research institutions that were already trained on very large datasets (e.g, Stable Diffusion), and then fine tune them on their own data.

For example, artist Lev Pereulkov fine-tuned the Stable Diffusion model 2.1 using 40 paintings by well-known "non-conformist" artists who worked in USSR starting in the 1960s (Erik Bulatov, Ilya Kabakov, etc). Pereulkov's image series Artificial Experiments 1–10 (2023) created with this custom net, is an original piece of art that captures the artistic characteristics of these artists as well as their unique surreal and ludicrous semantics without repeating closely any of their existing works. Instead, their "DNAs" captured by the net enable new meanings and visual concepts.



Lev Pereulkov, Artificial Experiments 1–10, 2022. Three images from the series of 10 shared on Instagram.

Most of the millions of everyday people and creative professionals who employ generative media tools use them as is, and don't fine them further. This may change in the future as the techniques networks using our own data may become easier to use. But regardless of these specifics, all newly created cultural artifacts produced by trained nets have a common logic.

Unlike traditional drawings, sculptures, and paintings, generative media artifacts are not created from scratch. They are also not the result of capturing some sort of sensory phenomenon, such as photos, videos, or sound recordings. They are instead built from *a large archive of other media artifacts. This generative mechanism links generative media to earlier art genres and processes.*

We can compare it to film editing, which first appears around 1898, or even earlier composite photography, which was popular in the nineteenth century. We can also consider specific artworks that are especially relevant, such as experimental collage film A Movie (Bruce Conner, 1958) or many Nam June Park installations that feature edited fragments of TV footage.

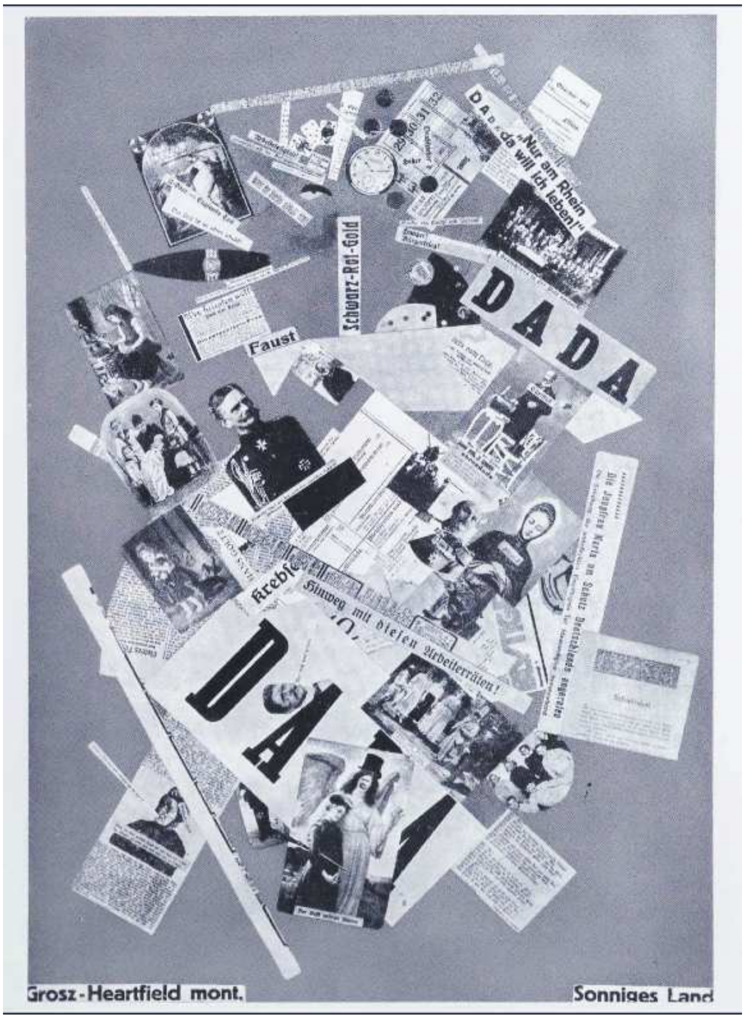
Seeing projects like *Unsupervised* or *artificial experiments 1-10* in the context of this media creation method and its historical variations will help us understand this and many other AI artworks as art objects engaged in dialogues with art from the past, rather than as purely technological novelties or works of entertainment.

I see many relevant moments and periods when I scan the history of art, visual culture, and media for other prominent uses of this procedure. They are relevant to the current generative media not only because artists working at these times used the procedure, but also because the reason for this use was consistent in all cases. *A new accumulation and accessibility of masses of cultural artifacts led artists to create new forms of art driven from these accumulations.* Let me describe a few of these examples.

Net and digital artists created a number of works in the late 1990s and early 2000s in response to the new and rapidly expanding universe of the world wide web. Health Bunting's _readme (1998), for example, is a web page containing the text of an article about the artist, with each word linked to an existing web domain corresponding to that word. Mark Napier's Shredder 1.0 (also 1998) presents a dynamic montage of elements that comprise numerous websites - images, texts, HTML code, and links.

Going further back in time, we find a broad cultural paradigm that was also a reaction to the accumulation of historical art and culture artifacts in easily accessible media collections. This is paradigm is known as "post-modernism." Post-modern artists and designers frequently used bricolage and created works consisting of quotations and references to art from the past, rejecting modernism's focus on novelty and breaking with the past.

While there are many possible explanations for the emergence of the post-modern paradigm in the 1960s and 1980s, one is relevant to our discussion. The accumulation of earlier art and media artifacts in structured and accessible collections such as slides libraries, film archives, art history textbooks with many photos of the artworks, and other formats - where different historical periods, movements, and creators were positioned together - inspired artists to begin creating bricolages from such references as well as extensively quoting them.



Photomontage by John Heartfield, 1919.

What about "modernism" in the 1910s and 1920s? While the overall emphasis was on originality and novelty, one of the procedures it developed in search of novelty was direct quotations from the vast universe of contemporary visual media that was rapidly expanding at the time. Large headings, for example, and the inclusion of photos and maps made newspapers more visually impactful; new visually oriented magazines, such as *Vogue* and *Times*, were also launched in 1913 and 1923, respectively; and of course, a new medium of cinema continued to develop.

In response to this visual intensification of mass culture, in the early 1910s Georges Braque and Pablo Picasso began incorporating actual newspaper, poster, wallpaper, and fabric fragments into their paintings. A few years later, John Heartfield, George Grosz, Hannah Hoch, Aleksandr Rodchenko, and a handful of other artists began to develop photo-collage techniques. Photo-collage became another method of creating new media artifacts from existing mass media images.

Contemporary artworks that employ neural networks trained on cultural databases, such as *Unsupervised* or *artificial experiments 1-10*, continue a long tradition of creating new art from *accumulations of images and other media*. In this way, these works of art keep opening up new possibilities for art and its techniques, particularly those of what I referred to as earlier "database art (see my article Database as a Symbolic Form, 1998). The introduction of new methods for *reading cultural databases and creating new narratives from them* is part of this expansion.

Thus, *Unsupervised* neither creates collages from existing images, as did modernist artists of the 1920s, nor quotes them extensively, as did postmodern artists of the 1980s. Instead, the group trains a neural network to extract patterns from tens of thousands of MoMA's artworks. The trained net then generates new images that share the same patterns but don't look like any specific paintings. Throughout the course of the animation, we travel through the space of these patterns (e.g., "latent space"), exploring various regions of the universe of contemporary art. (For a more details about GAN net training methods used by Refik Anadol Studio, see "Creating Art with Generative Adversarial Network: Refik Anadol's Walt Disney Concert Hall Dreams," 2022).

Pereulkov's Artificial Experiments 1–10 use a different technique to generate new images from an existing image database. He chose only forty paintings by artists who share key characteristics. They developed their oppositional art in late communist society (USSR, 1960s-1980s). They also lived in the same visual culture. In my memories, this society was dominated by two colors: grey (representing the monotony of urban life) and the red of propaganda.

In addition, Pereulkov chose paintings that share something else: "I chose, as a rule, paintings that conceptually relate in some way to the canvas - or to the space on it. I obtained the painting "New Accordion" from Kabakov, which features paper applications on top of the canvas" (my personal communication with Pereulkov, 04/16/2023). Pereulkov also crafted custom text descriptions of each painting used for fine-tuning the Stable Diffusion model. To teach the model the specific visual languages of the chosen artists, he added terms such as "thick strokes," "red lighting," "blue background," and "flat circles" to these descriptions.

Clearly, each of these steps represents a conceptual and aesthetic decision. In other words, the key to the success of *Artificial Experiments 1–10* is the creation of such a database. This work demonstrates how fine-tuning an existing neural network that was trained on billions of image and text pairs (such as Stable Diffusion) can make this network follow artists' ideas; the biases and noise of such a massive network can be overcome and minimized, and do not need to dominate our own imagination.

4. From Representation to Prediction

Historically, humans created images of existing or imagined scenes by a number of methods, from manual drawing to 3D CG (see below for explanation of the methods). With AI generative media, a fundamentally new method emerges. Computers use large datasets of existing representations in various media to predict new images (still and animated).

One can certainly propose different historical paths leading to visual generative media today, or divide one historical timeline into different stages. Here is one such possible trajectory:

1. Creating representations manually (e.g. drawing with variety of instruments, carving, etc). More mechanical stages and parts were sometimes carried out by human assistants typically training in their teacher's studio – so there is already some delegation of functions.
2. Creating manually but using assistive devices (e.g. perspective machines, camera lucida). From *hands* to *hands + device*. Now some functions are delegated to mechanical and optical devices.
3. Photography, x-ray, video, volumetric capture, remote sensing, photogrammetry. From *using hands* to *recording information using machines*. From *human assistants* to *machine assistants*.
4. 3D CG. You define a 3d model in a computer and use algorithms that simulate effects of light sources, shadows, fog, transparency, translucency, natural textures, depth of field, motion blur, etc. From *recording* to *simulation*.

5. Generative AI. Using media datasets to predict still and moving images. From *simulation* to *prediction*.

"Prediction" is the actual term often used by AI researchers in their publications describing visual generative media methods. So, while this term can be used figuratively and evocatively, this is also what actually happens scientifically when you use image generative tools. When working with a text-to-image AI-model, the neural network attempts to predict the images that correspond best to your text input. I am certainly not suggesting that using all other already accepted terms such as 'generative media' is inappropriate. But if we want to better understand the difference between AI visual media synthesis methods and other representational methods developed in human history, employing the concept of 'prediction' and thus referring to these AI systems as 'predictive media' captures this difference well.

5. Media Translations

There are several methods for creating 'AI media'. One method transforms human media input while retaining the same media type. Text entered by the user, for example, can be summarized, rewritten, expanded, and so on. The output, like the input, is a text. Alternatively, in the image-to-image generation method, one or more input images are used to generate new images.

However, there is another path that is equally intriguing from the historical and theoretical perspectives. 'AI media' can be created by automatically 'translating' content between media types. This is what happens, for example, when you are using Midjoiurney, Stable Diffusion or other AI image generator service and enter a text prompt, and AI generates one or more images in response. Text is 'translated' into an image.

Because this is not a literal one-to-one translation, I put the word 'translation' in quotes. Instead, input from one medium instructs a neural network to predict the appropriate output from another. Such input can also be said to be 'mapped' to some outputs in other media. Text is mapped into new styles of text, images, animation, video, 3D models, and music. The video is converted into 3D models or animation. Images are 'translated' into text, and so on. Text-to-image method translation is currently more advanced than others, but various forms will catch up eventually.

Translation (or mapping) between one media and another is not a new concept. Such translations were done manually throughout human history, often with artistic intent. Novels have been adapted into plays and films, comic books have been adapted into television series, a fictional or non-fictional text was illustrated with images, etc. Each of these translations was a deliberate cultural act requiring professional skills and knowledge of the appropriate media. Some of these translations can now be performed automatically on a massive scale thanks to artificial neural networks, becoming a new means of communication and culture creation. Of course, artistic adaptation of a novel into a film by a human team and automatic generation of visuals from novel text by a net is not the same thing, but for many more simple cases automatic media translation can work well. What was once a skilled artistic act is now a technological capability available to everyone. We can be sad about everything that might be lost as a result of the automation – and democratization – of this critical cultural operation: skills, something one might call 'deep artistic originality' or 'deep creativity', and so on. However, any such loss may be only temporary if the abilities of 'culture AI' are, for example, even further improved to generate more original content and understand context better.

Because the majority of people in our society can read and write in at least one language, text-to-another media methods are currently the most popular. They include text-to-image, text-to-animation, text-to-3D, and text-to-music models. These AI tools can be used by anyone who can write, or by using readily available translation software to create a prompt in any of the language these tools understand best at a given point. However, other media mappings can be equally interesting for professional creators. Throughout the course of human cultural history, various translations between media types have attracted attention. They include translations between video and music done by VJs in clubs; long literary narratives turned into movies and television series; texts illustrated with images in various media such as engravings; numbers turned into images (digital art); texts describing paintings (ekphrasis tradition, which began in Ancient Greece), mappings between sounds and colors (especially popular in modernist art); etc.

The continued development of AI models for mappings between all types of media, without privileging text, has the potential to be extremely fruitful, and I hope that more tools will be able to accomplish this. Such tools will be very useful both to professional artists and other creators alike. However, being an artist myself, I am not claiming that future 'culture AI' will be able to match, for example, innovative interpretations of Hamlet by avant-garde theatre directors such as Peter Brook or astonishing abstract films by Oscar Fishinger that explored musical and visual correspondences. It is sufficient that new media mapping AI tools stimulate our imagination, provide us with new ideas, and enable us to explore numerous variations of specific designs.

6. The Stereotypical and the Unique

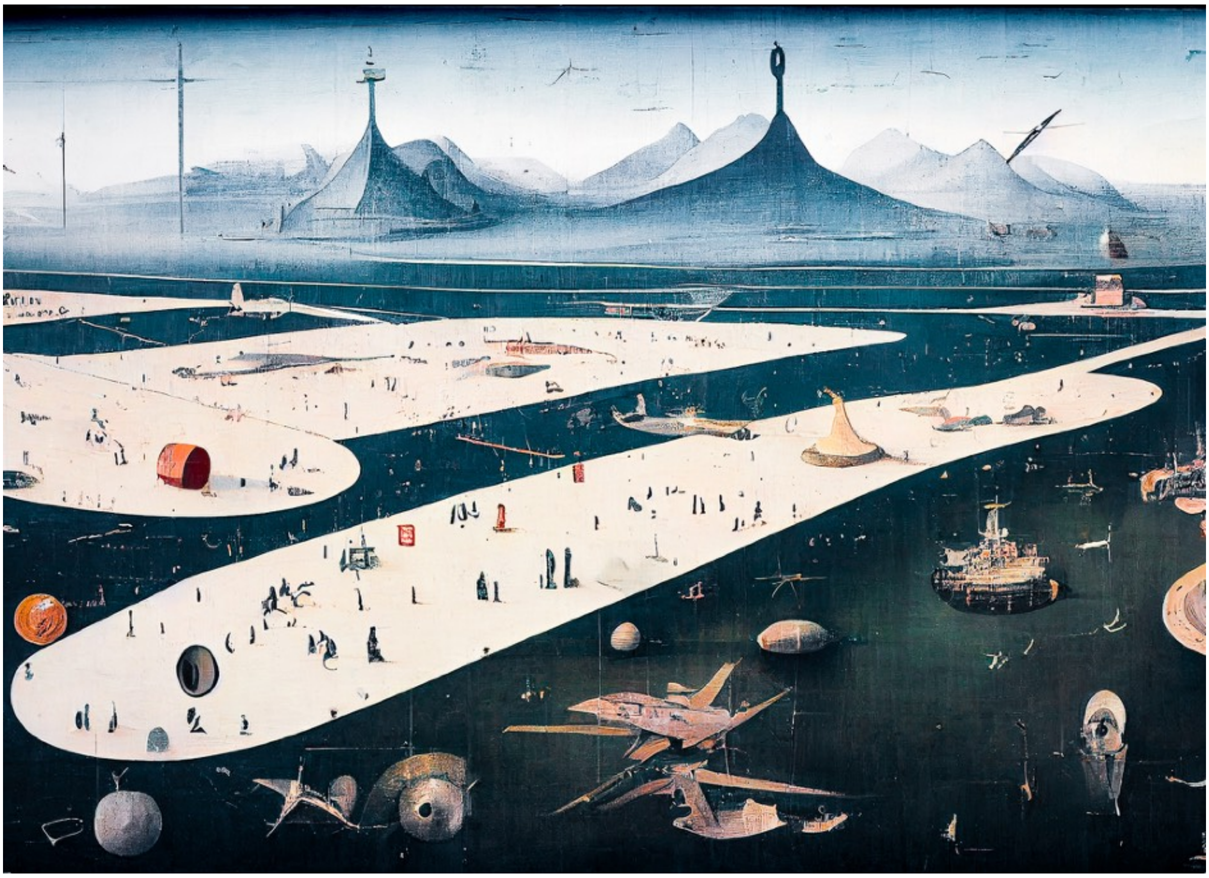
Both the modern human creation process and the predictive AI generative media process seem to function similarly. A neural network is trained using unstructured collections of cultural content, such as billions of images and their descriptions or trillions of web and book pages. The neural net learns associations between these artifacts' constituent parts (such as which words frequently appear next to one another) as well as their common patterns and structures. The trained net then uses these structures, patterns, and 'culture atoms' to create new artifacts when we ask it to. Depending on what we ask for, these AI-created artifacts might closely resemble what already exists or they might not.

Similarly, our life is an ongoing process of both supervised and unsupervised cultural training. We take art and art history courses, view websites, videos, magazines, and exhibition catalogs, visit museums, and travel in order to absorb new cultural information. And when we 'prompt' ourselves to make some new cultural artifacts, our own biological neural and networks (infinitely more complex than any AI nets to date) generate such artifacts based on what we've learned so far: general patterns we've observed, templates for making particular things, and often concrete parts of existing artifacts. In other words, our creations may contain both exact replicas of previously observed artifacts and new things that we represent using templates we have learned, such as golden ratio or use of complementary colors.

AI neural nets used for image generation frequently have a default 'house' style. This is the actual term used by MidJourney developers). If one does not specify a style explicitly, the AI will generate it using this 'default' aesthetic.

Examples generated in Midjourney version 4 using text prompt "morning sky."

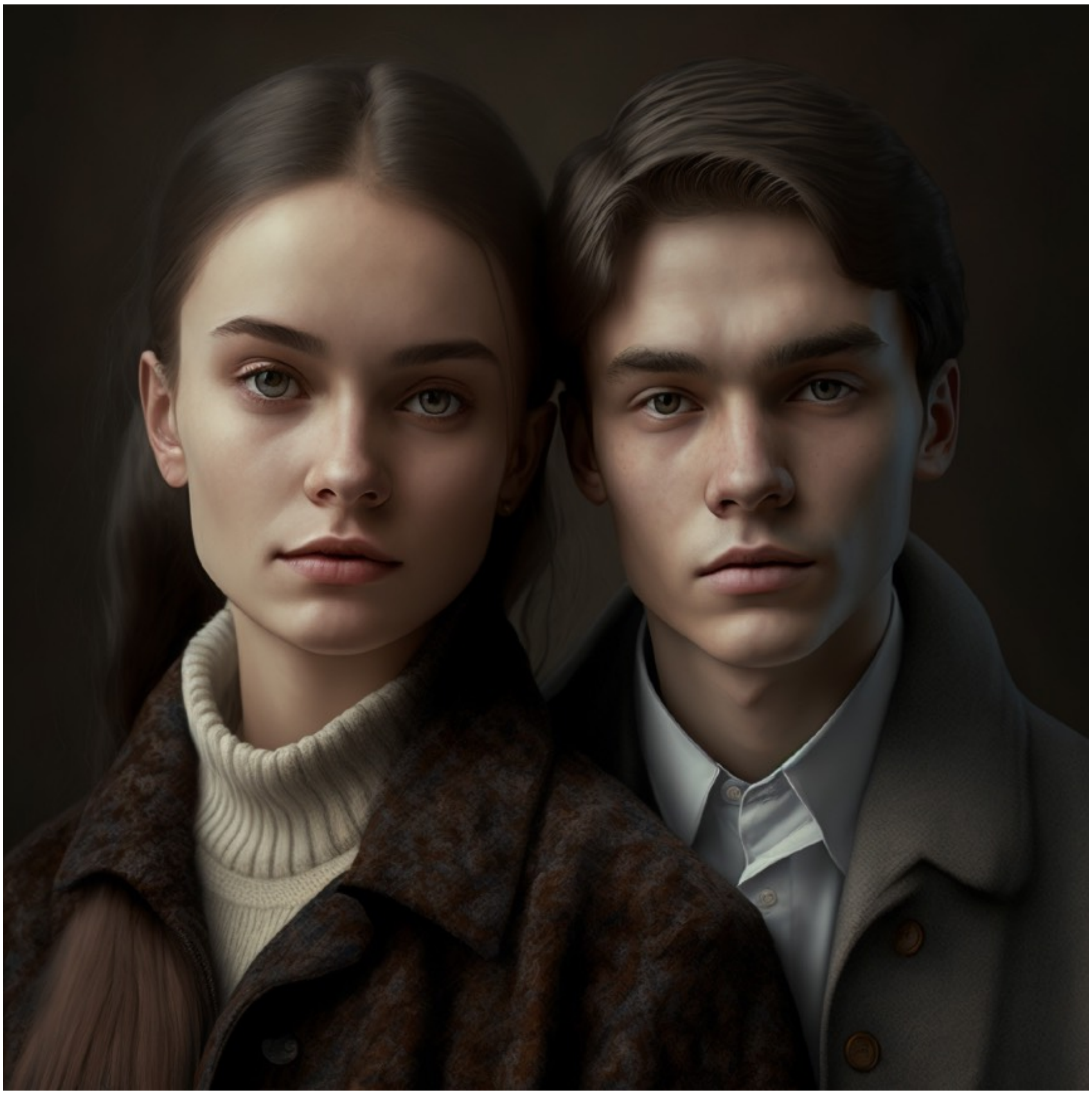
To steer away from this default, you need to add terms to your prompts specifying, a description of the medium, the kind of lighting, the colors and shading, or a phrase like "in the style of" followed by the name of a well-known artist, illustrator, photographer, fashion designer, or architect. Here are two examples of such prompts I made, and the images that Midjourney generated from these prompts. The terms used to define particular style characteristics are highlighted in red.



Prompt 1:

"giant future 1965 modern airport in Siberia made from water and ice, painted on large wood panel by Hieronymus Bosch, bright pastel colors with white highlights, 23f lens, very detailed --ar 4:3 --s 1250 —test"

(Image generated with Midjourney v3)



Prompt 2:

"Photo of two Russian high-school students, clear skin, very soft studio light, 50mm lens, monochrome, silver tones, high quality, ultra realistic --v 4 --q 2"

(Image generated with Midjourney v4)

This image also illustrates the point I am making later in the chapter: "AI frequently generates new media artifacts that are more stereotypical or idealized than what we intended."

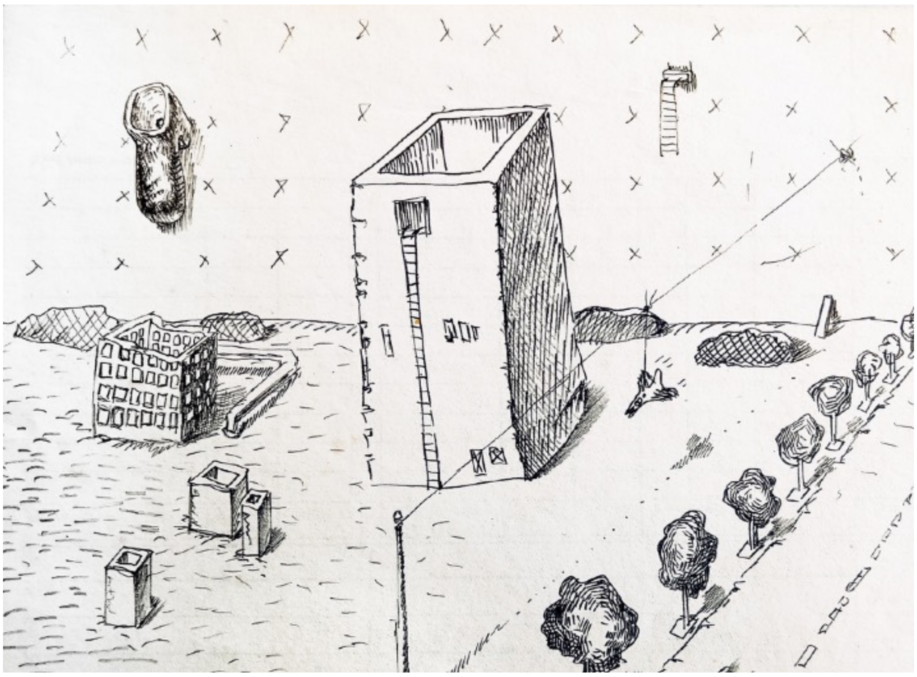
Because it *can simulate many thousands of already-existing aesthetics and styles and interpolate between them to create new hybrid*s, AI is more capable than any single human creator in this regard. However, at present, skilled and highly experienced human creators also have a significant advantage. Both humans and artificial intelligence are capable of imagining and representing nonexistent and existing objects and scenes alike. Yet, unlike AI image generators, human-made images can include very particular content, unique minuscule details, and distinctive aesthetics way that is currently beyond the capabilities of AI. In other words, today a large group of highly skilled and experienced illustrators, photographers, and designers can represent everything a trained neural net can do (although it will take much much longer), but *they can also create objects, compositions, or aesthetics that the neural net cannot do at this time. Equally importantly, they can picture unique objects, faces, compositions, and so on - as opposed to often more common-place or idealized versions generated by AI.*

What is the cause of this aesthetic and content gap between human and artificial creators? 'Cultural atoms', structures, and patterns in the training data that occur most frequently are very successfully learned during the process of training an artificial neural network. In the 'mind' of a neural net, they gain more importance. On the other hand, 'atoms' and structures that are rare in the training data or may only appear once are hardly learned or not even parsed at all. They do not enter the artificial culture universe learned by AI. Consequently, when we ask AI to synthesize them, it is unable to do so.

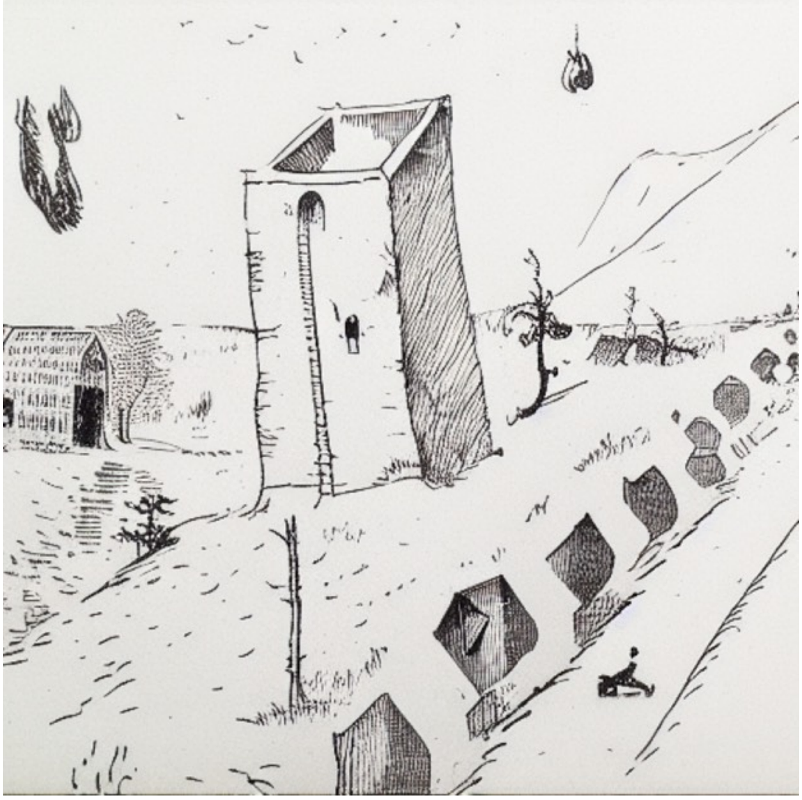
Due to this, text-to-image AIs such as Midjourney, Stable Diffusion or RunwayML are not currently able to generate drawings in my style, expand my drawings by adding newly generated parts, or replace specific portions of my drawings with new content drawn in my style (e.g, they can't perform useful "outpainting" or "inpainting" on the digital photos of my drawings.) Instead, these AI tools generate more generic objects than what I frequently draw or they produce something that is merely ambiguous yet uninteresting.

I am certainly not claiming that the style and the world shown in my drawings is completely unique. They are also a result of specific cultural encounters I had, things I observed, and things I noticed. But because they are uncommon (and thus unpredictable), AI finds it difficult to simulate them, at least without additional training using my drawings.

Lev Manovich, untitled drawing, pen on paper, 1981-1982.



One of my attempts to generate a new version of this image with Stable Diffusion AI, Fall 2022.



Here we encounter what I see as *the greatest obstacle creators face when using AI generative media:*

AI frequently generates new media artifacts that are more stereotypical or idealized than what we intended.

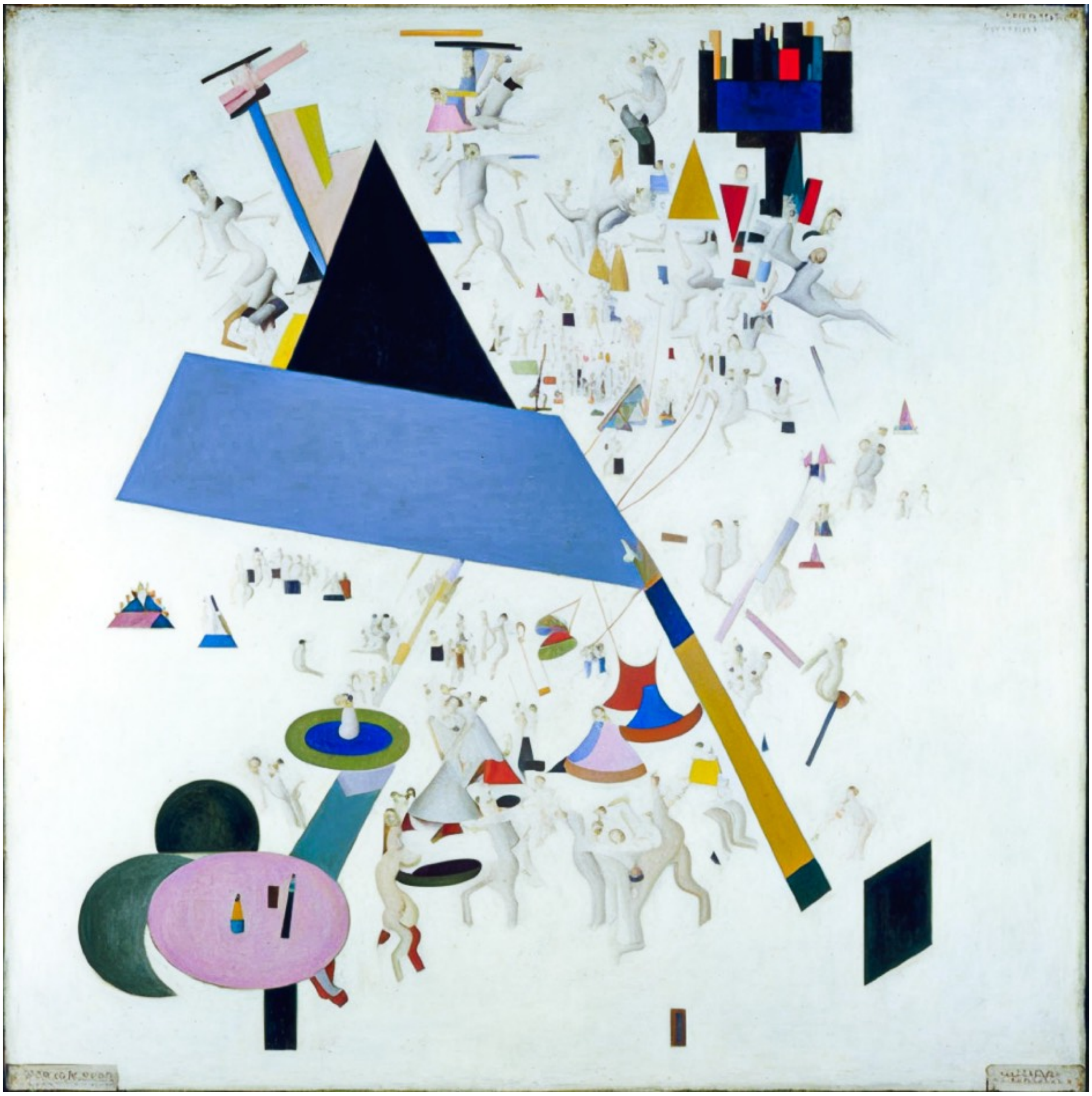
This can affect any image dimensions - elements of content, lighting, crosshatching, atmosphere, spatial structure, and details of 3D shapes, among others. Occasionally it is immediately apparent, in which case you can either attempt to correct it or disregard the results. Very often, however, *such 'substitutions' are so subtle that we cannot detect them* without extensive observation or, in some cases, the use of a computer to quantitatively analyze numerous images. In other words, new AI generative media models, much like the discipline of statistics since its inception in the 18th century and the field of data science since the end of the 2010s, deal well with frequently occurring items and patterns in the data but do not know what to do with the infrequent and uncommon. We can hope that AI researchers will be able to solve this problem in the future, but it seems so fundamental that we should not anticipate a solution immediately.

7. Subject and Style

In the arts, the relationship between 'content' and 'form' has been extensively discussed and theorized. This brief section does not attempt to engage in all of these debates or to initiate discussions with all relevant theories. Instead, I would like to consider how these concepts play out in AI's 'generative culture'. However, instead of using content and form, I'll use a different pair of terms that are more common in AI research publications and online conversations between users: ***subject*** and ***style***.

At first glance, AI media tools appear capable of clearly distinguishing between the subject and style of any given representation. In text-to-image models, for instance, you can generate countless images of the same subject. Adding the names of specific artists, media, materials, and art historical periods is all that is required for the same subject to be represented differently to match these references. Photoshop filters began to separate subject and style as soon in the 1990s, but AI generative media tools are more capable. For instance, if you specify "oil painting" in your prompt, simulated brushstrokes will vary in size and direction across a generated image based on the objects depicted. AI media tools appear to 'understand' the semantics of the representation as opposed to earlier filters that simply applied the same transformation to each image region regardless of its content. For instance, when I used "a painting by Malevich" and "a painting by Bosch" in the same prompt, Midjourney generated an image of space that contained Malevich-like abstract shapes as well as many small human and animal figures like in popular Bosch paintings that were properly scaled for perspective.

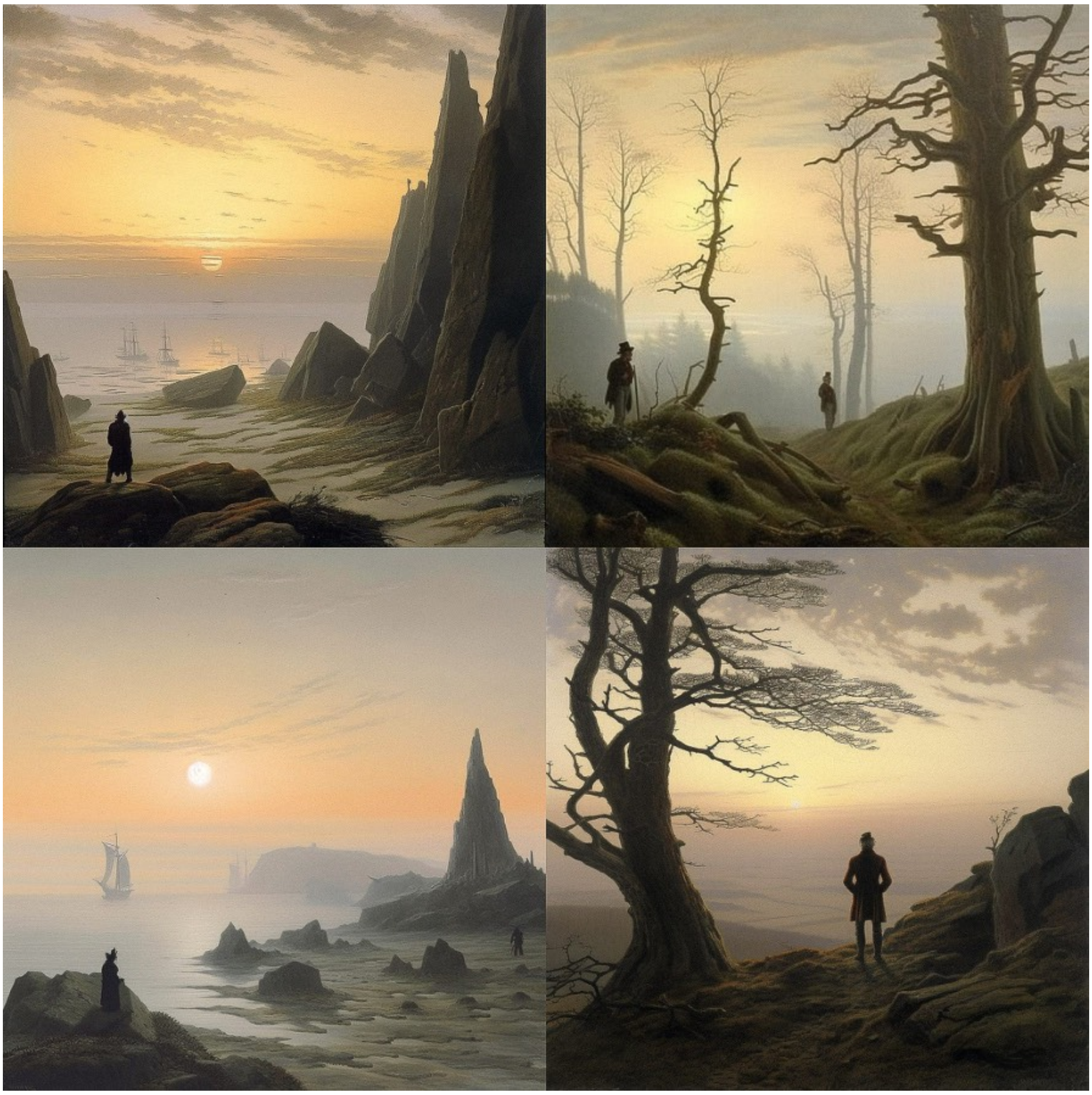
Image generated in Midjourney using prompt "painting by Malevich and Bosch," Fall 2022.



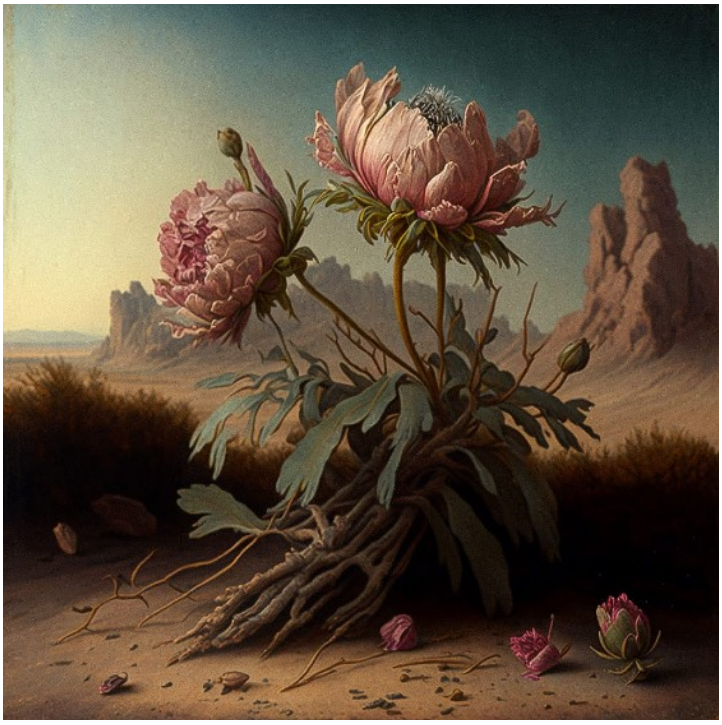
AI tools routinely add content to an image that I did not specify in my text prompt in addition to representing what I requested. This frequently occurs when the prompt includes “in the style of” or “by” followed by the name of a renowned visual artist or photographer. In one experiment, I used the same prompt with the Midjourney AI image tool 148 times, each time adding the name of a different photographer. The subject in the prompt remained mostly the same – an empty landscape with some buildings, a road, and electric poles with wires stretching into the horizon. Sometimes adding a photographer’s name had no effect on the elements of a generated image that fit our intuitive concept of style, such as contrast, perspective, and atmosphere. But every now and again, Midjourney also modified the image content. For example, when well-known works by a particular photographer feature human figures in specific poses, the tool would occasionally add such figures to my photographs. (Like Malevich and Bosch, they were transformed to fit the spatial composition of the landscape rather than mechanically duplicated.) Midjourney has also sometimes changed the content of my image to correspond to a historical period when a well-known photographer created his most well-known photographs.

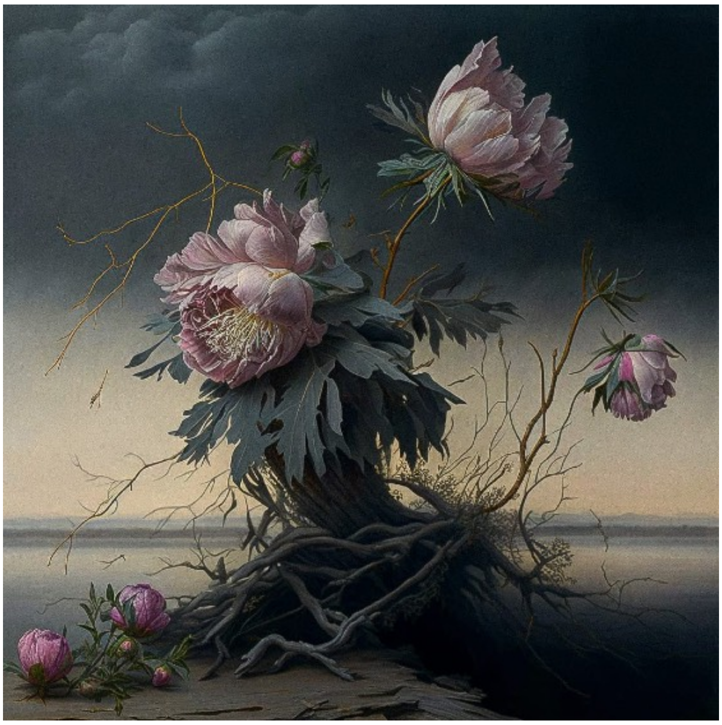
According to my observations, when we ask Midjourney or a similar tool to create an image in the style of a specific artist, and the subject we describe in the prompt is related to the artist’s typical subjects, the results can be very successful. However, when the subject of our prompt and the imagery of this artist are very different, ‘rendering’ the subject in this style frequently fails.

Using prompt “by Caspar David Friedrich --v 5” in Midjourney generates images that capture the artist’s style sufficiently well. Source: https://www.midlibrary.io/styles/caspar-david-friedrich.



Using prompt “decaying peonies by Caspar David Friedrich” in Midjourney generates images that simulate important features of artist’s style such as combinations of cool colors and dramatic atmosphere. But in other ways, generated images depart significantly from the artist’s style. The types of lines, rendering of details, and symmetrical compositions in these AI images would never appear in actual Friedrich’s paintings. AI can also often insert some generic looking objects, such as the rock formations in the upper right corner of first image.





To summarize, in order to successfully simulate a given visual style using current AI tools, you may need to change the content you intended to represent. *Not every subject can be rendered successfully and satisfyingly in any style*. Additionally, AI can often successfully learn some features of artist's style but not others.

These observations, I believe, complicates the binary opposition between the concepts of 'content' and 'style'. For some artists, AI can extract at least some aspects of their style from examples of their work and then apply them to different types of content. But for other artists, it seems, their style and content cannot be separated. (This is only my initial observation to be developed further after doing more experiments in Midjourney.)

For me, these kinds of observations and reflections are one of the most important reasons for using new media technologies like AI generative media and learning how they work. Of course, as a practicing artist and art theorist, I had been thinking about the relationships between subject and style (or content and form) for a long time but being able to conduct systematic experiments like the one I described brings new ideas and allows us to *look back at cultural history and art in new ways*.