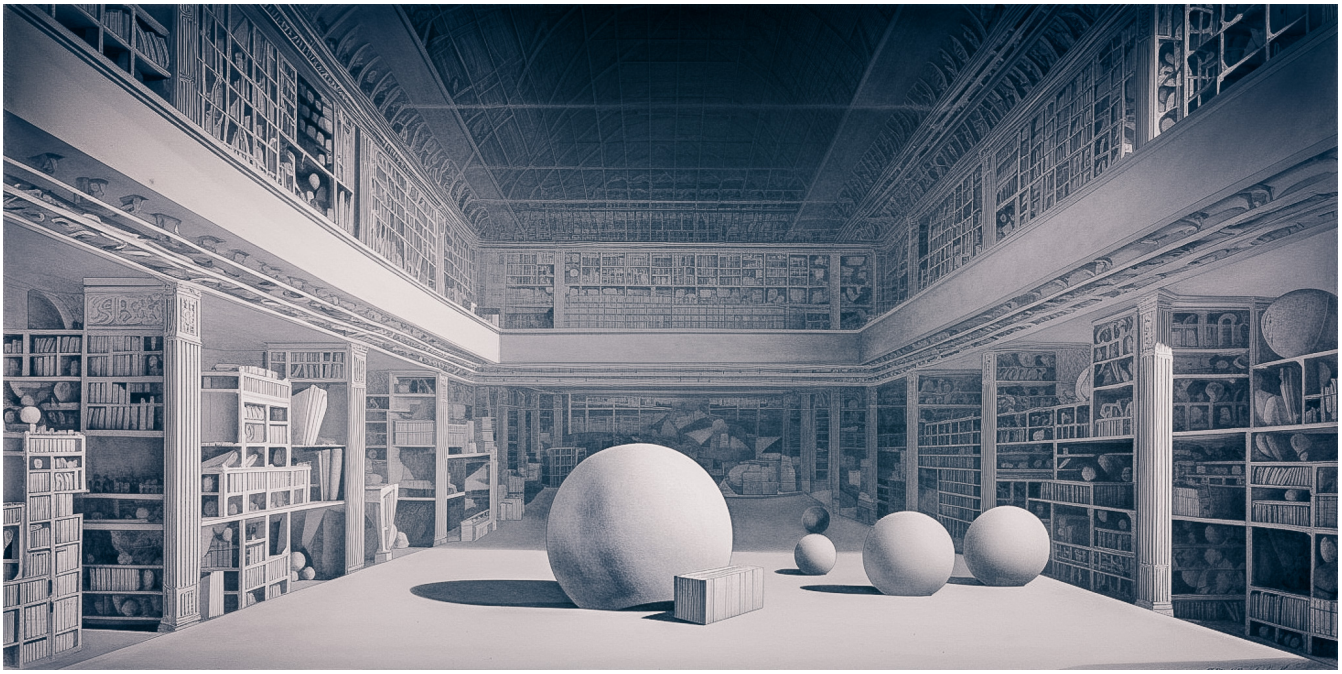
Lev Manovich

Separate and Reassemble: Generative AI Through the Lens of Art and Media Histories

Chapter 7 of *Artificial Aesthetics* by Lev Manovich and Emanuele Arielli. Published March 26, 2024. Other chapters: http://manovich.net/index.php/projects/artificial-aesthetics



Lev Manovich, from *Architecture for a Universal Library* series, generated in Midjourney and edited in Lightroom, 2023.

The use of terms:
In this chapter, the terms *generative media*, *synthetic media*, or *generative AI* refer to the process of synthesizing media objects with artificial neural networks. The examples of such objects are include text, voice, music, 3D models, datasets, and computer code. The terms *generative image, AI Image* or *visual AI* refer to specially synthesized visual objects. These objects can be still images that imitate the appearance and structure of all types of visual media, from photographs to drawings, and also moving images that imitate appearances of animation and video.

Separate and Reassemble

AI image represents a further logical evolution of the process that begins with digital media algorithms in the 1970s and continues in the following decades. The first computer paint programs were created in the 1970s, but could not yet simulate different paint types, brushes, and textured surfaces like canvas.[1] But in the 1990s, software such as Coral Painter (1991–) started to offer these features.[2] Similarly, the first 3D computer graphics algorithms for rendering solid shapes, Gouraud shading (1971) and Phong shading (1973), couldn't yet simulate the looks of different materials. Later, in the 1970s and 1980s, computer graphics researchers created numerous algorithms to simulate the appearance of various materials and textures, such as cloth, hair, and skin, as well as shadows, transparency, translucency, depth of field, lens flares, motion blur, reflections, water, smoke, fireworks, explosions, and other natural phenomena and cinematography techniques and effects.

Simulating many of these phenomena and techniques requires multiple separate algorithms that were developed over time. Thus, we find distinct sessions devoted to such algorithms with names like Volumes and Materials, Fluid Simulation, or Cloth and Shells in the annual proceedings of SIGGRAPH, the main conference in CG field.[3] As an example, the paper "Predicting Loose-Fitting Garment Deformations Using Bone-Driven Motion Networks" presented in 2023 conference describes "a learning algorithm that uses bone-driven motion networks to predict the deformation of loose-fitting garment meshes at interactive rates." Another conference paper "Rendering Iridescent Rock Dove Neck Feathers" describes a new approach for modeling and rendering bird feathers; and so on.

In my 1992 article "Assembling Reality: Myths of Computer Graphics" I have analyzed this fundamental aspect of computer graphics, explaining that "synthetic photorealism is

---

[1] On the history of early paint programs, see Alvy Ray Smith, *Digital Paint Systems: An Anecdotal and Historical Overview*, https://ohiostate.pressbooks.pub/app/uploads/sites/45/2017/09/paint.pdf, and also his *A Biography of the Pixel* (The MIT Press, 2021).

[2] https://en.wikipedia.org/wiki/Corel_Painter.

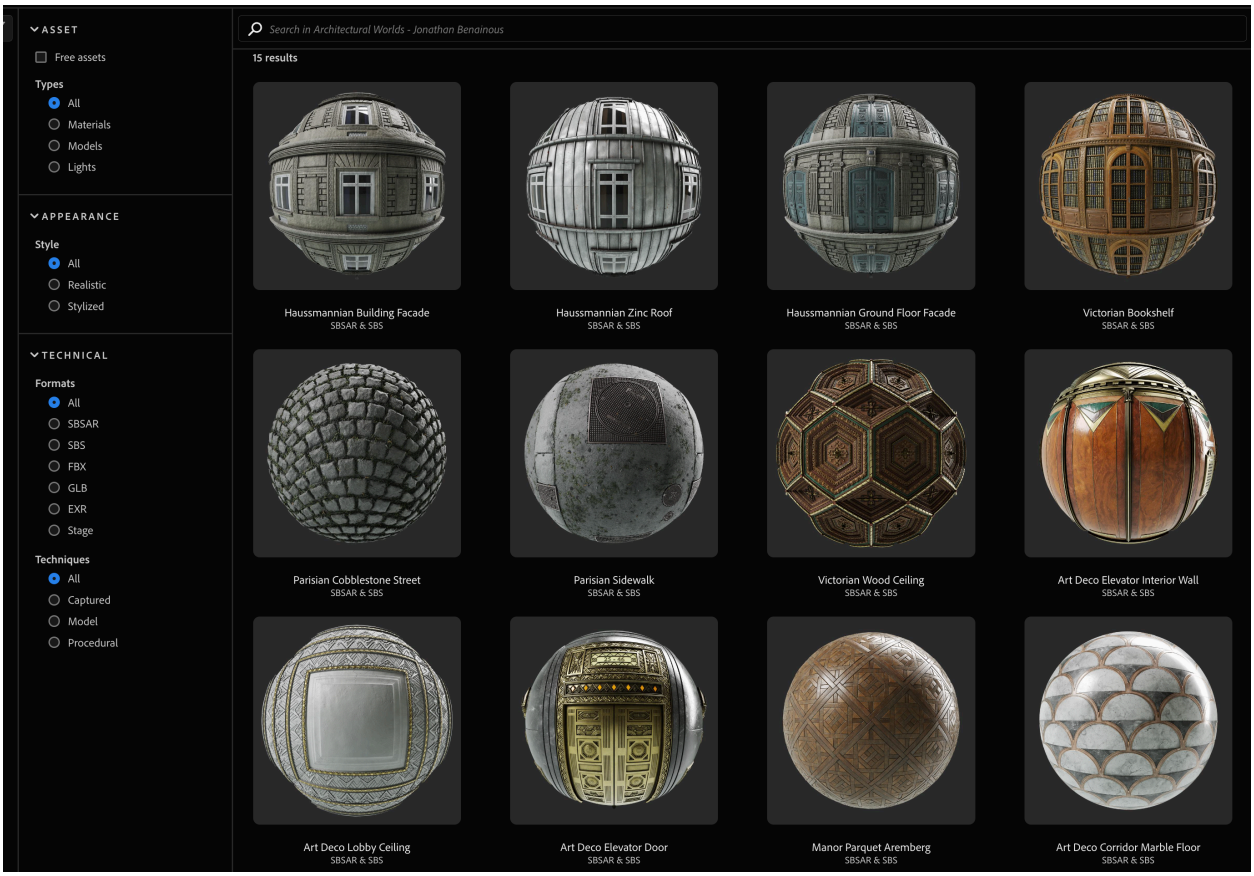
[3] See *SIGGRAPH '22: ACM SIGGRAPH 2022 Conference Proceedings*, https://dl.acm.org/doi/proceedings/10.1145/3528233.

fundamentally different from the realism of the optical media, being partial and uneven, rather than analog":

> Digital recreation of any object involves solving three separate problems: the representation of an object's shape, the effects of light, and the pattern of movement. To have a general solution for each problem requires the exact simulation of underlying physical properties and processes. This is impossible because of the extreme mathematical complexity… In practice, computer graphics researchers have resorted to solving particular local cases, developing a number of unrelated models for simulation of some kinds of shapes, materials and movements.[4]

In other words, 3D CG takes the world which we see apart, separating objects' shapes, materials, light reflections, textures, movements and behaviors. During rendering, the effects of multiple algorithms simulating all these aspects are combined together. Thus, *visual representations created using CG are discrete and modular, rather than continuous and "monistic."* This is one of the most important characteristics of CG medium, distinguishing it from lens-based optical image media.

A few from the thousands of materials available in Adobe 3D content creation software.



This logic of separation and recombination also defines next stage of digital media: PC software for media creation and editing. Following its initial release in 1990, Photoshop gradually began to include simulated effects and techniques from various artistic mediums, ranging from darkroom photography to oil painting, within a single program. These effects can be combined in a single digital image. Music software similarly allows users to combine many simulated instruments and multiple effects such as reverb and echo in one composition. Word processing and desktop publishing software separate the physical process of print composition into its basic parts that also can be now recombined - for example, you can take any font and arbitrary change its size or generate your own font.[5]

All of these media software capabilities were first proposed in the 1970s and later realized in the 1980s and 1990s, eventually becoming ubiquitous. AI generative media follows the same logic, although its underlying technical implementation is different. During training, neural networks learn visual patterns characteristic of hundreds of different types of art media, lighting techniques and effects from history of photography and cinematography, and visual signatures of many thousands of historical and contemporary artists, architects, fashion designers and other creators. A reference website Midlibrary currently lists 367 "artistic techniques" that Midjourney AI image generator tool can reliably simulate according to the tests conducted by this website team.[6] They range from "albumen print" and "anaglyph" to "wood carving" and "wireframe rendering. "

Importantly, a user can include references to multiple techniques and/or multiple creators in a single prompt, potentially generating new types of media effects that did not exist before. Here are examples of such prompts from my own experiments:

Using multiple artists in one prompt: "18th century very big and detailed panoramic etching showing landscape in the style of *Michael Kaluta, Kawanabe Kyosai, Pieter Bruegel the Elder*, insane detail, cinematic"

Using multiple artistic media in one prompt: "18th century futuristic infinite museum storage space with art objected on the shelves, snow fall inside the space and fog, wide angle view looking down, 7pm soft evening light, detailed intricate *drawing and etching* with very fine shading, subtle nuanced sombre *color pencils* and fine *pens"*

---

[5] For the detailed analysis of media software and its conceptual origins, see Lev Manovich, *Software Takes Command* (Bloomsbury Academic, 2013).

[6] https://midlibrary.io/, accessed February 25, 2024.

A screenshot from midlibrary.io showing a few of artistic techniques, art genres, and “styles” of painters, illustrators, architects, photographers, and fashion designers that Midjourney can simulate. At the moment of this writing, the library contains close to 5000 such references. (Captured March 24, 2024.)

The pioneering digital media theorist of 1990s and 2000s William J. Mitchell[7] called this key characteristic of digital media "separate and recombine." In his 1995 book *City of Bits*, he described this process in relation to urban planning:

> Classical architects of the eighteenth and nineteenth centuries handled the task of putting spaces together by creating hierarchies of great and small spaces around axial, symmetrical circulation systems connected to grand, formal entries and public open spaces…functionalist modernists of the twentieth century have often derived their less regular layouts directly from empirically established requirements of adjacency and proximity among the necessary spatial elements. But when telecommunication through lickety-split bits on the infobahn supplements or replaces movement of bodies along circulation paths, and when telepresence substitutes for face-to-face contact among the participants in activities, the spatial linkages that we have come to expect are loosened. The constituent elements of hitherto tightly packaged architectural and urban compositions can begin to float free from one another, and they can potentially relocate and recombine according to new logics.[8]

Mitchell's lectures in the 2000s expanded on this formulation, demonstrating how the logic of separation and recombination can be seen in digital media in a variety of ways. Generative AI continues the same logic. A neural network extracts elements and structures from hundreds of millions or billions of images in its training set. They include distinct color palettes, compositions, lighting effects, artifacts of historical photography processes, and so on. When you ask AI image tool to generate new images with specified visual attributes, it does its best to combine (or more precisely, *interpolate* between) appropriate art patterns and effects.

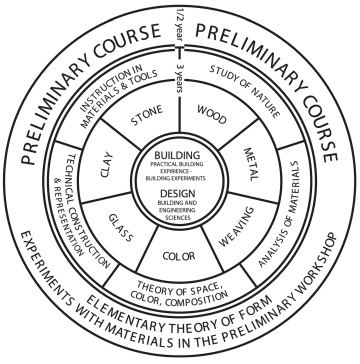
No human historian, theorist or practitioners of visual art, photography, cinema, or design were ever able to describe all such patterns. In the early 20th century pioneering art historians Aby Warburg and Erwin Panofsky developed the study of iconology. Warburg defines this concept as visual motives that (re)appear in various civilizations and media. Panofsky used it somewhat differently, referring to symbols and motifs that have existed throughout the history of art.

---

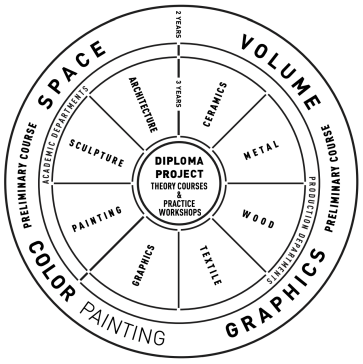
[7] See https://mitpress.mit.edu/author/william-j-mitchell-2911/.

[8] William J. Mitchell, *City of Bits: Space, Place, and the Infobahn*, revised edition (The MIT Press, 1996).

During the same period visual artists and architects disassembled visual arts in a different way, breaking down an image into its basic components and dimensions such as points, lines, planes, two-dimensional forms, color, space, texture, pattern, balance, and equilibrium, among others. While this project of methodical dismantling and creation of new visual languages from these components is central to modernist art and its many -isms, it finds its most methodical development in the curricula of two new schools of art and design. VKhUTEMAS in Moscow (1920-1930) and Bauhaus in Germany (1919-1933) introduced around the same time their Basic Course where students where taught how to systematically work with all these elements and dimensions. Instead of drawing from life, painting portraits or making historical compositions, now students started training by completing exercises with image primitives such as basic shapes, forms, and colors. At VKhUTEMAS the Basic Course was created in 1920 by Rodchenko, Popova, Ekster, Vesnin and other faculty from painting, architecture, and other school's areas. In its first iteration it consisted from a number of workshops such as "Discipline of Synchronized Shapes and Colors," Plane, Color and Spatial Design, "Graphic Construction on a Plane Surface" and "Color." It was further transformed during VKhUTEMAS existence. Eventually, three learning sequences were approved for all VKhUTEMAS students: Plane and Color, Volume and Space.[9] (The Basic Course at this school was more systematic and comprehensive than a similar course at Bauhaus; it was taught by many different faculty members and lasted two years.[10])



Bauhaus
Teaching Program Diagram, W. Gropius, 1922.

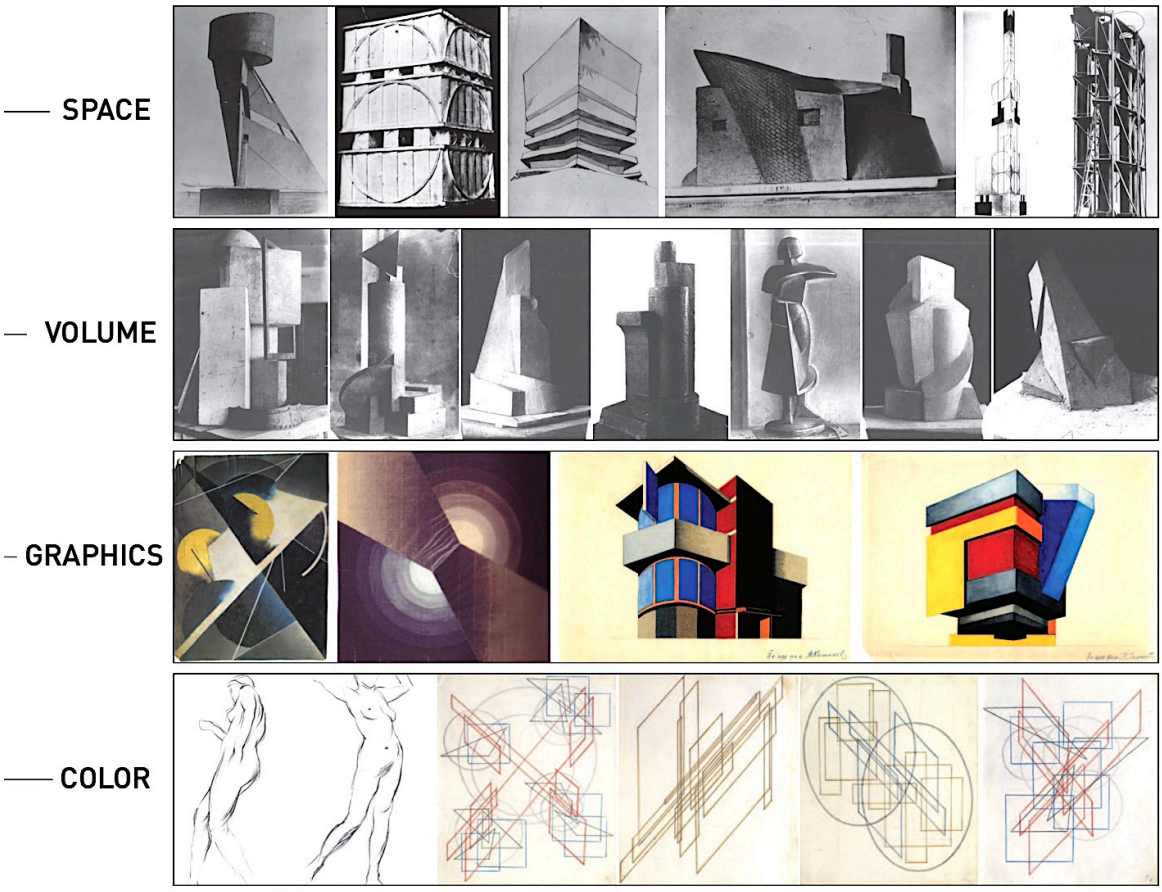


VKhUTEMAS
Teaching Program Diagram, 1923.

---

[9] https://www.vkhutemas.ru/en/structure-eng/faculties-eng/main-course/, accessed February 26, 2024. See also Anna Bokov, *Avant-Garde as Method: Vkhutemas and the Pedagogy of Space, 1920–1930* (Park Books, 2021).

[10] Note that VKhUTEMAS was ten times larger then Bauhaus, with 100 faculty and 5000 students during ten years of its existence versus only 500 students at Bauhaus.

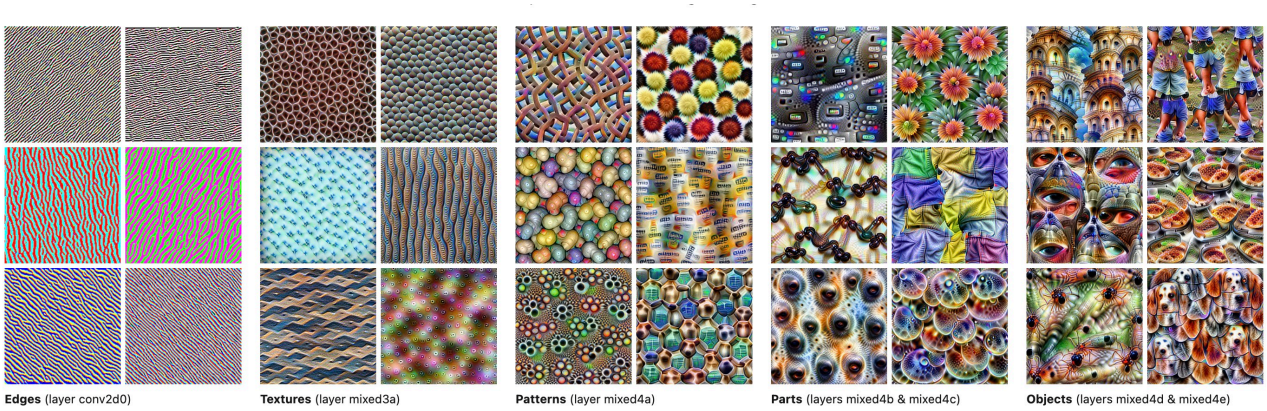


Previous page: the structures of courses in Bauhaus and VKhUTEMAS. Both curricula begun with the basic course (the outer ring). This page: examples of student exercises at VKhUTEMAS. From Anna Bokova, VKhUTEMAS Training, 2014, http://www.avantgardesculpture.com/downloads/VKhUTEMAS-Handout.pdf.

In a certain sense, *generative AI models can be said to continue these programs of decomposition and analysis of visual arts that begun in the early twentieth century*. Artificial intelligence algorithms extract patterns (or "features") from training data. However, at the moment, we can't look at billions of parameters in a gigantic generative network and get neat catalog of all the patterns the network learnt.[11] In the 2010s when

[11] See for example, Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, Robin Rombach, "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis,"arxiv.org, 4 Jul 2023, https://arxiv.org/abs/2307.01952.

neural networks were simpler and smaller, scientists were able to visualize what their neurons learn. For example, the following visualization shows the features learned by a network trained to recognize objects in photographs. A network first learns how to recognize basic features before progressing to object recognition. (Unfortunately, the architecture of generative networks that synthesize images prevents us from "looking inside" these networks and visualizing them in the same manner.)[12]



Google Research visualization showing the features learned by progressive layers of a network trained for image recognition. The first layers learn basic features such as edges and textures (see previous page), and later layers learn the appearance of objects parts and whole objects. Source: https://distill.pub/2017/feature-visualization/.

I want to conclude with a relevant quote from my 2018 book *AI Aesthetics*.[13] While at that time deep neural networks were mostly used for media classification and recommendations, with generative AI revolution still four years away, the analysis I developed in the book section called "AI as a Culture Theorist" has become even more relevant today:

> [There is] a crucial difference between an "AI culture theorist" and a human theorist/ historian. The latter comes up with explicit principles that describe how a cultural area function…a neural net can be trained to distinguish between works of

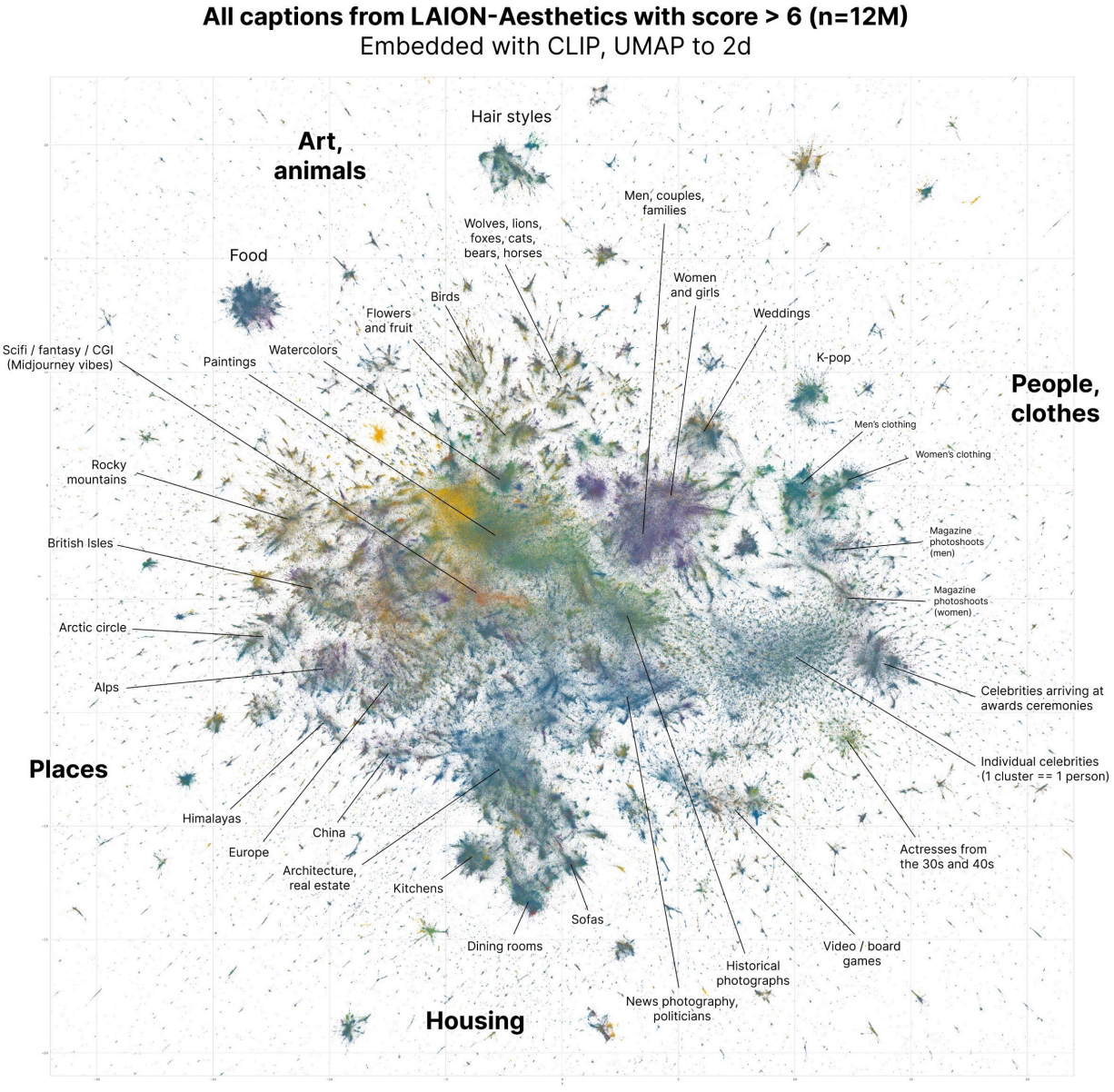
[12] For the overview of available deep networks visualization methods, see "How to Visualize Deep Learning Models," https://neptune.ai/blog/deep-learning-visualization, 14 November, 2023.

[13] Lev Manovich, AI Aesthetics (Strelka Press, 2018), http://manovich.net/index.php/projects/ai-aesthetics.

> different artists, fashion designers, or film directors. And it can also generate new objects in the same style. But often we don't know what exactly the computer has learned…Will the expanding use of machine learning to create new cultural objects make explicit the patterns in many existing cultural fields that we may not be aware of?

This theoretical potential to me is one of the most interesting and valuable thing about generative AI - but we will have to wait to see if it maybe realized in the future.

David McLure, visualization of 12M captions from LAION-Aesthetics dataset with aesthetic score > 6 (2022). LAION-Aesthetics is a part of 5.6 billion captioned images dataset used to train popular Stable Diffusion AI image generation model. This visualization gives us some idea about content of images in training data - but not the artistic techniques and styles model extracts from images.



Visual AI and Media Accumulation

Visual AI is the fourth significant *data <-> knowledge* effect of the web - a global accumulation of networked hyperlinked cultural content that began to grow quickly after 1993. Although people have been sharing texts and images on the internet since the 1970s, this process picked up speed after 1993, when the first visual browser, Mosaic, was introduced on January 23 of that year.

I have observed several repercussions of the growth of information on the web over the next 30 years. If we wish to situate the development of Visual AI in the early 2020s in this timeline, here are four such effects. Certainly others can be also named, so this is only one list of techno-cultural developments technologies enabled by the web I am particularly interested in:

1. The first effect is the switch from categorical, hierarchical and structured organization of information (exemplified by library catalogs and early web directories) to search engines in the late 1990s. There was so much content that organizing it in conventional ways was no longer practical, and search become the new default. Note that *web search is based on a prediction of what will be most relevant to the user* as opposed to giving you a precise and definite answer. Note that generative AI is also predictive - it predicts possible text, images, animation or music in response to your question or prompt. The regime of absolute certainty, i.e a truth vs a lie typical for human civilization is replaced by predictions, as statistics becomes foundation of human sciences in the 20th century, and data science and AI in recent decades.

2. The second major effect is the rise in popularity of data visualization during the 2000s. The field comes into its own around 2005. As a part of this development, the new field "artistic data visualization" develops in the same decade, along with other new cultural fields: data art and data design. (In our lab we created *Phototrails*, *Selfiecity* and *On Broadway* in 2012–2014. These were first interactive visualizations of millions of Instagram images.) If search attempts to find the most relevant items in the giant data universe, visualization tries to show parts of this universe in one image, revealing patterns and connections.

3. The third effect is the emergence of "data science" as the master discipline of the new big data era at the end of the 2000s. While many methods employed in data science

have already been available for decades, the rapid increase in unstructured data in the 2000s motivated the development of a separate data science field—the key new profession of the data society. My own version of this stage is "cultural analytics," the idea I came up with in 2005 and worked on for the next 15 years in our lab. Our main method was data visualization, but now applied to large media collections of photos, video, film, manga, magazine covers, Instagram images, etc. I named this method *media visualization*. (See projects at http://lab.culturalanalytics.info).

4. The next, but certainly not the last, effect of the growth of online visual digital content is Generative AI which becomes popular in early 2020s. Dalle-e is released in 2020, MidJourney in 2022, ChatGPT and Photoshop generative fill in 2023, and hundreds of other tools exist today. (A bit earlier around 2017, a particular AI method for media generation called GAN became already popular with digital artists.)

(It is relevant to mention that both Visual AI and Generative AI in general builds on 20 years of work, with the first relevant papers published in 2001. The key idea is to use web content universe as a source of data for ML, without labeling it, already appears in the research paper published around that time.)

Let's see what kind of pattern is established by these four effects. Search is the first method to deal with the new scale of content on the web. Data science focuses on finding patterns, relations, clusters, and outliers in big data, and also predicting future data. Data visualization tries to summarize datasets visually. And now Generative AI explores "big content" in yet another way, generating new content which combines many patterns from existing media.

To put this differently - Generative AI synthesizes new content that has statistical properties similar to existing content. But it's not a copy of what already exists. AI generates new content (texts, images, animation, 3D models, music, singing, etc.) by interpolating between existing points in the latent space. This space contains numerous patterns and structures extracted by artificial networks from billions of image-text pairs, trillions of text pages, and other large collections of existing human cultural artifacts. AI predicts what could exist between these points in space of patterns. For example, it can predict a "painting" made by artists A, B, C, using techniques D and E, with content F, G and E, with mood, colors M-N, proportion W, composition K, etc.

Note that the three earlier developments all approach big data by summarizing it. Web search reduces billions of web pages to the top results. Dara vis reduces it to a diagram.

Data science reduces it by using summary statistics, cluster analysis, regression or latent space projection. But Visual AI is doing something new. It also first reduces big data during learning and then generates new data points.

One way to sum up all this is to say that we we moved from probabilistic search to probabilistic media generation: 1999 to 2022. But certainly generative AI is not the last effect of the existence of web data; others will be likely emerge in the future.

Compression and Generation

> "If everybody speaks in the same way, everybody is voiceless…this is why poetry is playing such a role in culture. *Poet is the most individual of all speakers*." Yuri Lotman, lecture "Circles and Communities," *Conversations about Culture*, 1988.

Both human and machine cognition rely on *compression*. We perceive and understand the world through categories and types. Human art also uses compression. But there is a fundamental difference between how compression works in the arts and in AI. While artworks often depict characters, symbols, or scenes that condense human experiences, they also frequently contain many concrete and distinct details. This combination of general and concrete, predictable and unique is especially important for modern arts (19th-21st c). Modern artists compresses human world and experience into patterns, structures and types - but they also typically add very particular, rare and unexpected unexpected to these general patterns.

In contrast, when we train AI models, the training data is also compressed, and the particular and unique are omitted. Extracting patterns from the data in machine learning involves eliminating outliers and many unique details and only selecting most frequently appearing associations, characteristics and structures.

This kind of compression is the fundamental characteristics of generative AI. The same also holds for machine learning and statistics in general. When data is summarized, classified or used to predict future data, the most common is preserved and the rare is abandoned.

Given this, *can we expect AI to create artworks with enough concrete, unique, and subtle details? If all the most unique (i.e. rare) information is not preserved during training, where will it appear when we generate new artifacts?*

This is, in my opinion, a more interesting and relevant question than the one everyone asks: can AI be creative. As I discussed earlier in Chapter 4 AI and Myths of Creativity, the association of the arts with "creativity" is a recent notion that becomes established only in the Romantic period. For thousands of years, humans were creating artifacts that today we admired as the very best artworks ever created by our species - even though their authors did not have the goals of inventing anything new or being "creative." (In Christian tradition, there can be only one creator - God.) Instead, art was thought to serve other goals such as *imitation* - the concept central to understanding of arts from Plato and Aristotle until the second half of the 18th century.[14]

One way to think further about the difference between compression in art and in AI, data science and statistics is by considering the notion of *average*. In descriptive statistics we use average measures such as mean, median, or mode. The mean is the arithmetic average of a set of numbers. For example, the mean of a set of eight numbers (1,2,3,4,5,6,7,8) is 4. The mean captures average tendency of a data set - but does not preserve any of its details, or its specificity. We can come with infinite number of number sequences that will all have the same mean of 4. Note that they don't have to include "4" as one of the numbers. For example: (1,2,2,3,5,6,6,7) or (0,0,0,2,6,8,8,8), etc.

Furthermore, while in some cases most numbers are close to the mean (as in bell curve distributions), in others all numbers in the set can be far from the mean. So, in the first case, mean captures the overall pattern of a sequence, but in the second, it misrepresents it.

***

In contrast to statistical measures of central tendency, literary characters created by best realist writers in the 19th century - such as Balzac, Flaubert, Zola, Tolstoy - combine the general and the specific. They are not statistical abstractions of social classes or types of people retaining only the features which common to all people of this or that type - for example, this person is hard working and optimistic, this person tends to dream but not

[14] For discussions of imitation in relation to AI, see Mario Carpo, *Beyond Digital* (The MIT Press, 2023), and also his "A short but believable history of the digital turn in architecture," https://www.e-flux.com/architecture/chronograms/528659/a-short-but-believable-history-of-the-digital-turn-in-architecture/, March 2023.

act, etc. instead, they are presented as real concrete and unique individuals. In other words, in such literary characters, the general and the particular, the typical and unique are combined together.

Balzac who today is recognized as the the pioneer of realism movement in literature was explicit about the importance of concrete details. He wrote: "'the author firmly believes that details alone will henceforth determine the merit of works."[15] Balzac meticulously researched the places depicted in his novels, traveling to remote locations and comparing notes from multiple visits. In these novels, intricate details about locations can sometimes take up up to twenty pages. And while his characters represent a distinct set of societal types - the enticing mistress, the noble soldier, the rascal, and so on - they also depicted as particular individual. In short, Balzac was able to strike a balance between the individual's uniqueness and the portrayal of the type.

In the Introduction *The Human Comedy* (1842), Balzac contends that literary creation and scientific investigation are closely related activities. This alignment of realist method in literature with empirical science is very telling, but let's remember that the end goals of the two are different. *Science aims to produce only generalizations* in the forms of models, explanations and predictions of natural or social world. In contrast, the arts appeal to both our intellect and our senses. The sensory dimension of the arts are central to aesthetic theories by key thinkers such as Edmund Burke, Friedrich Nietzsche, Clive Bell and Susan Sontag, among others.

Here we can recall that Plato was very critical of visual arts precisely because they represent concrete reality. Plato saw the physical world as a mere shadow or replica of the true, ideal world of Forms (or Ideas), the most real and unchangeable parts of reality that can only be reached via rational contemplation. And this is why he considered visual arts (and other kinds of artistic imitation) as doubly removed from reality. However, if Plato could have been more perceptive, he would have noticed that sculptors in ancient Greece of his time were not only imitating the visible world or striving to compress it to reveal ideal Forms. Instead, in contrast to many other ancient cultures, Ancient Greek artists developed a special aesthetics that combined idealization and realism.

***

---

[15] For sources of this and other quotes from Balzac, see https://en.wikipedia.org/wiki/Honor%C3%A9_de_Balzac.

Here is another crucial difference between the arts and statistics. The former developed and practiced many different ways to compress phenomena. There is not one formula for summarization that fits arts made in all times, cultures and styles. For example, sculptures, tiles, reliefs, seals or paintings from Sumer, Babylon, Assyria and other ancient civilizations don't represent gods and kings with obsessive photorealism of 19th century artists such as Jean-Auguste-Dominique Ingres (1780-1867) or Ivan Shishkin (1832 - 1898). Instead, the details of human bodies and faces appear schematized, i.e. (to use our terms) "compressed," But the techniques of such compression vary enormously. The outlines of bodies, the proportions of body parts, the features of faces, the shapes of clothes are represented in variety of ways. What details are kept and what is simplified, how this simplification is carried, what is presented realistically and what is exaggerated can change from culture to culture.

To summarize this discussion, we can say that while both human mimetic arts and statistics, data science, and AI use compression, for arts its only an option and not a requirement. Moreover, an artwork can have both general patterns and concrete non-aggregated details. And last but not learn, artworks can employ a variety of ways to create their patterns.

This does not mean that in practice generative AI tools are always inferior to very accomplished human creators because they can't always generate enough small and specific details. Sometimes they can. However, as many users have noticed, often they do struggle to produce sufficient variety of unique details. This may not be a problem if I aim for schematic and aggregated aesthetics - i.e., if I want only archetypes. Ancient, classical and modern arts give us plenty of examples of great artworks which use such approach. However, pursuing different aesthetics that combine general and concrete and have a high degree of individualization, such as faces in Jan van Eyck paintings, descriptions of feelings, thoughts, people, and places in Proust, or architecture details in Antoni GaudÃ's buildings, can be challenging at present.

Alexander Deineka (1899-1969), *Tekstile Workers*, 1927, oil on canvas, 171 **x** 195 cm.

In his famous painting, Deineka skillfully blends general and concrete, abstract and unique. The factory's machinery and architecture have been reduced to almost abstract geometric patterns. The rendering of bodies emphasizes their similarity, while faces retain their uniqueness. The level of detail in the faces appears unexpected and almost exaggerated when compared to the machinery's deliberately very regular patterns. These are just a few examples of the artist's selective and variable "compression" techniques used in this and other future works.

The Aesthetics of Fragments

In one of my image series created with Midjourney, you human figures and interior spaces with shelves filled with endless objects. They remind me of a book library, a warehouse, natural history museum display, European still-live paintings… Some of the objects can be recognized, but others look like fragments. In another image series I have been making for over a year now, we see young artists inside their spacious art studios. Here also all surfaces - floor, walls, and even ceiling - are also covered with dense textures and patchworks of lines and shapes. It is not clear if they are two or three-dimensional. Do these lines belong to the paintings covering the walls or do they exits in actual space?.. Some of these details are purely abstract. Others appear to suggest something definite, some objects, shapes and meanings from our human world.

What are these tiny fragments? What are these undefined unnamed objects and shapes filling endless shelves, covering the floor, or growing to fill the space?

Yes, these are “fragments” - but of what?

Lev Manovich, this and previous page: from *Drawing Rooms* series, generated in Midjourney and edited in Lightroom, 2023-2024.

In an art or archeology museum we see fragments of ancient civilizations. Pieces of vases, glasses, plates, and also small tools, statues etc. in other words, these are parts of single concrete objects from 4000, 2000, 1500 years ago. But AI "fragments" have diffident ontology. During AI model training, patterns" from hundreds of millions of images are extracted and distributes them across trillions of connections.

In this training process, digital materially of images is further virtualized, evaporated, diffused - but still preserved. And the generated "fragments" you see in my images are like scents, invisible movement of slight wind, or periodic movements of the ocean edge leaving traces on the sand.

These are fragments of fragments, in other words. Deposits of already broken forms. More fragmented then the 18th century ruins admired by painters and visitors in Italy on Grand Tour. They are not like "glitches." And they are not noise noise of telecommunication networks theorized by Claude Shannon.

They are artifacts of one possible AI aesthetics - its distributed knowledge and distributed vision. The future internet protocol invented over 60 years ago in late 1950s by Paul Baran suggested braking whole messages into pockets in order to transmit them over distant networks more reliably. Breaking something into random parts paradoxically assured its survival. (And this how Internet still works today.)

Paul Baron published first description of the "pocket switching" concept in 1960 when he was working for RAND in the US. Around the same time, Alexey Ivakhnenko and Valentin Lapa, two Soviet mathematicians working in Ukraine, invented another fundamental method for using fragments of a message in a productive fashion. Their invention was the first deep neural network - which eventually led to contemporary massive networks that also used in Generative AI.[16]

Breaking cultural artifacts during machine learning into fragments (such as image pixels and parts of the words) and then processing these fragments in stages eventually allows these networks to acquire knowledge that can produce synthetic text, images, music,

---

[16] See Juergen Schmidhuber, *Annotated History of Modern AI and Deep Learning*, https://arxiv.org/abs/2212.11279.

spaces, code. In short, by breaking historical human culture into fragments we get our new "generative culture."

The wonderfully coherent texts "written" by ChatGPT are predicted one word at a time. This blind "language computer" can't see far ahead, beyond one word. But somehow, one word is associated with another, and the second is associated with the third. Pulling these threads forward gives us poems, fiction stories, job cover letters, textbook chapters, computer programs…

And when you select a single one-pixel wide column in an image, another blind "image computer" continues this line of pixels (I am referring to generative AI 'in-painting' technique). A single line gives rise to endless magnificent new worlds. Their coherence and familiarity contradicts a seemingly random RGB values of a single column of pixels that I chose.

In other words, we were assuming that we were looking at something without any meaning, any sense, any value - because we randomly selected a line in a drawing or a photograph. So for our human vision, it was random. But we did not consider the fact that this line was part of a larger area with coherent patterns - be it houses, a road, smiling faces, clouds, or any other possible subject. So the blind computer actually looked at all this and predicted what else can be nearby. (This is also how Photoshop's "generative fill" and "generative expand" tools work.)

The history of digital media, and also the history of all human civilization is in serious need of a *theory of fragments*. A taxonomy that will establish that there are dozens of such species. Certainly others have thought about this already - again I recall 18th century European aesthetics of ruins as one example.

Echoing the progress of modern science in its pursuit for the elements of matter such as molecules, atoms, and elementary particles, 19th century European artists begin to earnestly fragment the visible world. Impressionists broke it into into separate colorful brushstrokes; Georges Seurat and Paul Signac relied on mechanical looking irregular dots. Another crucial next step was taken by Paul Cezanne around 1878. He declared that all shapes should be represented in terms of the cylinder, the sphere and the cone. But paradoxically, this seeming solidification of reality was simply a new and more extreme way to fragment it. While impressionists visualized fragments of our perception, Cezanne

fragmented the world itself. The table cloth in his paintings such as, for example, *The Card Players* (1892-1985), is no longer a single continuous piece of fabric. Instead, it is a set of fragments - a collection of flat planes oriented in multiple directions.

In the 1907-1932 the aesthetics of violent fragmentation came to the center of visual modernism. We see it everywhere: in collages of Picasso and Braque; cubism of Jean Metzinger, Albert Gleizes, Fernand Léger, Robert Delaunay, etc.; cubo-futurism of Natalia Goncharova, Aleksandra Ekster, Lyubov Popova, Malevich and others; photomontages of Hannah Hoch, Lazlo Moholy-Nagy, Aleksander Rodchenko; film montage of Vertov and Eisenstein; and even cultural history montage of Walter Benjamin's unfinished *The Arcades Project*.

The latter had led to many interpretations that suggested more theories of cultural fragmentation. For example, in his *Theory of the Avant-Garde* (1984), art theorist Peter Bürger defines Benjamin's understanding of allegory as "a four-part schema that involves: first, the isolation and removal of a fragment from its context; second, the combination of fragments to create new meanings; third, the interpretation of the allegorist's gaze as melancholic – as one that draws 'life' out of the objects assembled; and finally, an understanding of allegory as a representation of history in decline rather than progress."[17]

In time, AI theorists may suggest equally interesting theories of network fragmentation which forms the basis of generative media. Perhaps we will even see giant visualizations of unimaginable resolution showing all patterns extracted by deep networks from their image databases. But as an artist who was always invested in digital aesthetics, I am simply happy to be generating my own private idiosyncratic image spaces with these barely visible "fragments" of something which does not fully exist. The fragments of one possible AI aesthetics.

The history of human culture is one of slow forgetting and very rare, almost statically impossible (one in hundred thousands? One in a million?…) remembering. Famous artists, writers, politicians, influencers commanding everybody's attention in their day disappear from historical memory and are absent from our records. And for the lucky few who are

---

[17] Jane Rendell, "Fragment of the imagination: assembling new narratives from old," *The Architectural Review*, 14 July 2021, https://www.architectural-review.com/essays/keynote/fragment-of-the-imagination.

remembered, it's only a few things. An artist work over sixty years is reduced to a few iconic images. History compression is brutal and uncompromising.

The mechanism of fragmentation and selective recombination and synthesis offered now by generative AI is a slightly less brutal. Of course, to qualify for possible remembering when deep neural network are being trained, something has to be lucky to have been digitized to begin with and/or end on the web. One painting in a small museum in the town that had no tourists visiting for the last four years got lucky because it is in the background of the selfie taken by a local high school student and her boyfriend visiting over the weekend. But all other paintings in this museum were not in the frame, so no neural network will learn about them.

Yes we get fragments anyway in this new cultural memory system - but at least they are more numerous than what more restrictive human memory and more brutal print culture allowed for before. This new more forgiving and less brutal process of forgetting and resurrection is perhaps the best aspect of AI aesthetics.

Lev Manovich, from *Architecture for a Universal Library* series, generated in Midjourney and edited in Lightroom, 2023.

A Letter to a Young Artist

> "I have completed the construction of my burrow and it seems to be successful." "... [T]he most beautiful thing about my burrow is the stillness. Of course, that is deceptive. At any moment it may be shattered and then all will be over. For the time being, however, the silence is with me." - Franz Kafka, *The Burrow,* 1924.

The key difference between me, a human, and generative AI: I am limited, but AI is unlimited. Yes, of course: it has significant limits now, in practice. But it advances fast, and what it can already do today is beyond what we could have imagined a year ago. During this one year (11/2022 - 11/2023), we got ChatGPT, Dalle-3, AI functions added to Photoshop, Midjourney v5… Instead of dwelling on what AI can't do at this particular moment, it is safer to assume that what it "can" will only multiply.

Because of how human skills, learning and memory works, I have limitations. I can't draw in hundreds of styles of other artists or effortlessly combine them together. I don't have knowledge of the immense *museum without walls* distributed over the web and museum databases. But AI can. And it will only get better.

I can't simply sit down and start writing summaries of numerous topics in the history of culture. AI can. I can't instantly make hours of music that mixes the languages of different composers and map them into new instruments. AI can.

"I can't… but AI can." (Endless other examples can be added.)

So why make art now? And what art will still be meaningful to make?

What is interesting about human art now is our limits - and obsessions. Our inability to instantly think and paint exactly like any one of the millions of artists who lived. Our inability to quickly change. The way I walk, talk, my habits. My constraints. This is what makes me human as opposed to an AI. The latter will continue to evolve. But human evolution does not work on the same scale.

Note that this is not about simulating my idiosyncrasies and thus making AI "more human." Yes we can do it, but that's not interesting. It is like taking a Boing 777 around the

block to get groceries. Its forcing super-humans to act like humans, and this is a banal and weak strategy.

And there is another crucial point to make. *What makes art "human" is not our intentions, plans, ideas or meanings. For over 100 years, modern artists did their best to remove all this from their art making.* If you give AI a direction, it can perfectly simulate ideas, plans and meanings. So this is not relevant.

The only relevant thing is our limitations. Our inability to compete with the superhuman. With the web, with search engines, with recommendation engines, with huge databases, with machine learning algorithms, with Generative AI - and other super-human computer technologies to come.

Therefore, "human artists making art with AI tools" is a meaningless idea. You want to collaborate with Gods? A mortal "collaborating" with Apollo, Athena, Hephaestus, Hermes, Zeus?

Instead, nurture your limitations. Be extremely limited—not unlimited. Don't be "creative." Forget the meaningless idea that AI will help us "expand our creativity."

Instead, work within constraints—the ones you already—or the ones you can make on purpose. White on white. Black on black. This is the right direction. Instead of a vast surface of "endless possibilities," concentrate in a single spor and go as deeply as possible.

(Think like Morandi rather than like Picasso.)

Make a tiny hole in the vast surface of everything that was already created and everything that is still possible, and keep digging. When you get completely tired digging meters of wrong underground paths, get lost again and again, and want to give up, it means you are finally close to something. Keep going.

Because AI is so vast and endless in its knowledge and skills, you needed to work on the micro-scale. Very narrow. So narrow that AI can't quite get there. Through the needle eye. Only in this way can you compete with superhuman generative AI.

The artist needs to become a mole. And you need to be constantly stressed and worried because AI can discover your hole at any time and, in an instant, destroy all the underground pathways you have spent years making. But perhaps this stress, this endless anxiety, is the right motivation for making something original and authentic in the end.

Making your art in secret, knowing that you can be discovered and erased tomorrow by AI progress.

Lev Manovich, *In the Garden*, generated in Midjourney and edited in Lightroom, 2023.