

Emanuele Arielli

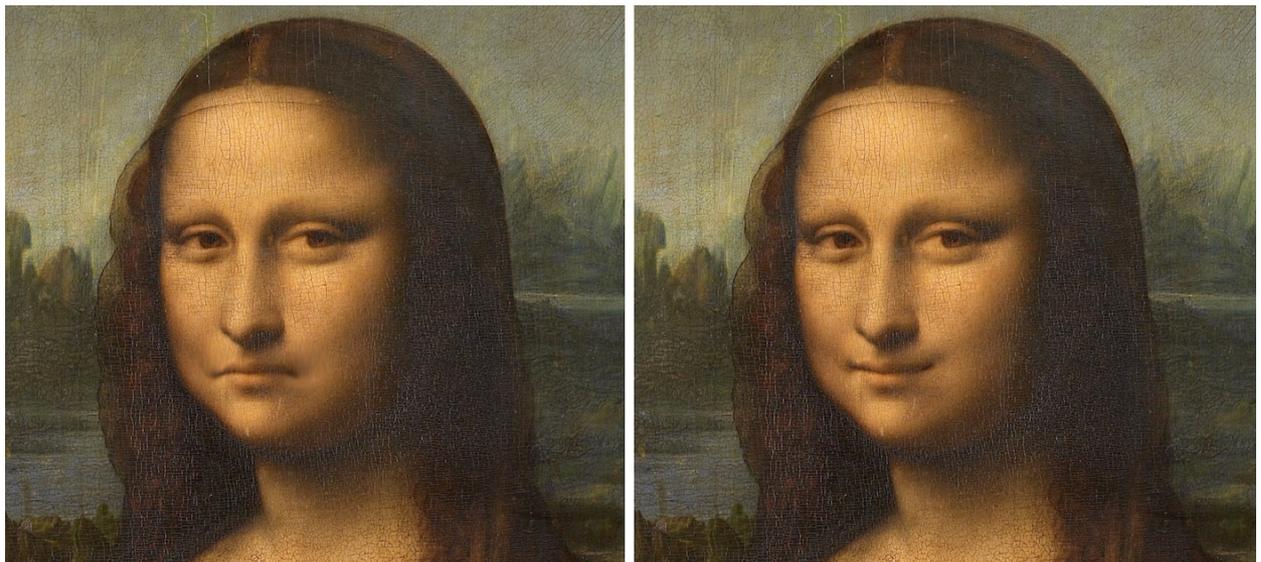
# Human Perception and The Artificial Gaze

Chapter 6 of *Artificial Aesthetics* by Lev Manovich and Emanuele Arielli

Other chapters: <http://manovich.net/index.php/projects/artificial-aesthetics-book>

Published January 15, 2024.

7,945 words.



“Sad” and “happy” Mona Lisa.

Mona Lisa face (left) appears more sad than in the original version (right). The facial expressions in two images varies overall; this difference is particularly pronounced in the mouth and eyes, which convey a dissimilar appearance. But if you cover the mouth in both images, you will see that all features of the face are actually identical, and that the sole alteration lies in the mouth’s corner. Essentially, we almost “hallucinate” an illusory expressive change that does not exist.

## The Innocent eye

In the 19th century, critic and historian John Ruskin, when discussing how one should enjoy a painting, famously stated: “The whole technical power of painting depends on our recovery of what may be called the innocence of the eye; that is to say, of a sort of childish perception of these flat stains of color, merely as such, without consciousness of what they signify, — as a blind man would see them if suddenly gifted with sight”<sup>1</sup>.

**If it were up to Ruskin, artificial systems analyzing images would have been the perfect viewers and critics of artworks, since they are “without consciousness of what they signify”.** In this view, machines, akin to children, would possess an unbiased and transparent “innocent eye”, able to see things as they are. But the question arises: do we ever perceive things in this manner at all?



Figure 1. What do you see see?

## **If it were up to Ruskin, artificial systems analyzing images would have been the perfect viewers and critics of artworks, since they are “without consciousness of what they signify”**

Let's examine the following image (Figure 1<sup>2</sup>): what do we see? To many, it may appear as a random arrangement of black and white shapes. But if we look at its original image (at the end of this chapter, Figure 11) and then return to this one, our perceptual impression changes drastically. We are now able to “read” it. One might wonder if the initial impression of random blots would be an example of “innocent” perception, or it is just the state of a “clueless eye”? And if that were the case, how would we be able to look “innocently” at things that are clearly recognizable?

What we recognize, what we know and what we expect can reshape our perception. The case of Figure 1 is akin to the experience of hearing a sentence in an unfamiliar foreign language and then, after acquiring proficiency, listening to it again. Initially perceived as a meaningless jumble of sounds, it is later understood as a structured sentence. Through language learning, we not only classify and recognize those sounds correctly but also experience a phenomenological and perceptual shift. Similarly, our past experiences, cultural habits, values, needs or personal attitudes also determine how we perceive things. One crucial point is that after we have learned to recognize a sound or an object, it becomes practically impossible to revert to a 'blank' and innocent perceptual state. I cannot ignore that sequence of sounds as a sentence in the language I am now acquainted with, and I cannot “unsee” the object I saw at the end of this chapter in Figure 1.

The process of learning categories unfolds in a similar manner. In early childhood, we come across various objects — dogs, cats, chairs, bicycles — without knowing what they are. As we learn to categorize them, often aided by verbal labels, our recognition of these objects changes. The first encounter with a cat is vastly different from later encounters when we recognize it as part of the broader category of *cats*. While sensory input constitutes the *bottom-up* aspects of perception, providing raw data from our

environment, our cognitive frameworks and the knowledge we've accumulated over act as *top-down* factors in perception.



You



Now again: what do you see here?



ChatGPT

This image is a visual illusion that can be seen in two ways. At first glance, it appears as a collection of black shapes against a white background. However, if one looks more closely, it becomes apparent that these shapes form the image of a panda. The black areas represent the typical markings of a panda—around the eyes, the ears, and the body—while the white areas represent the rest of the face and the fur. This kind of illusion demonstrates how our visual perception can shift between seeing abstract shapes and recognizing a familiar object.

Figure 2. Curiously, ChatGPT Large Language Model sees a panda in this image.

Computers are increasingly adept at performing human-like perceptual tasks. In some instances, they can even surpass human capabilities in detecting finer details and patterns, as evidenced in many impressive data processing applications, where machines are increasingly able to interpret X-rays images in medical diagnostics, identify faces even if they are masked, identify a car's model from an engine's noise, or evaluate the authenticity of a painting. However, and we might say fortunately, many human perceptual skills that we take for granted still constitute a challenge for artificial vision. This is particularly evident in the occasional failures of object detection in self-

driving cars. Despite these challenges, there is notable progression in this domain. Another critical consideration is how similar human and computer vision are or *should be*, that is, how well computer perception and AI analysis of sense-data are able to model human perceptual experience.

In 2015, a Google scientist unveiled a series of images produced by a Deep Dream convolutional neural network, also making the code publicly available. This release sparked significant interest, leading to widespread creation of similar images. These images, often eerie biomorphic landscapes filled with animal features emerging from ordinary scenes, demonstrated how the neural network identified specific shapes within its visual input. Essentially, the network's visual system is programmed to identify and emphasize certain elements such as, for example, dogs or eyes. It modifies the original image, aligning it with these predetermined patterns<sup>3</sup>. This process reveals the ongoing efforts of the machine to interpret images, molding them into the forms the program is designed to recognize. While these images might appear "hallucinatory" to the human observer, it is crucial to avoid taking too literally the claim that the machine "hallucinates" anything or that it subjectively "sees" these forms in the images it processes. Nevertheless, it represents a step toward the development of computer vision systems that integrate human-like mechanisms such as perceiving according to categories and expectations. By exploring how machines process and modify visual data, we gain insights into the intricate process of visual perception, both in artificial and human contexts.

Some might argue that for most AI applications, the ability of machines to mimic human sight or hearing is not relevant; what is important is their capacity to execute specific tasks and solve problems. Yet, this viewpoint might overlook key aspects regarding aesthetic applications. In scenarios where AI systems are tasked with generating content intended for human engagement, aiming to evoke aesthetic, emotional, and sensory responses, it becomes essential for those system to have a model of how humans perceive the world (see Chapter 1, "Studying subjects"). This task is not unattainable, yet it is more complex than merely devising systems that see the world "as it is". On the one hand, we perceive less than the full spectrum of available sensory data, as its complexity and our cognitive limitations compels us to filter out of certain elements. On the other hand, our perception extends beyond the sensory input, due to our ability to organize and interpret these data. This process enables us to extract deeper meanings and discern patterns that go beyond a mere aggregation of individual stimuli. Consider the so-called *Gestalt* laws in object recognition. An an example, these laws describe how we recognize a shoe in an image

like that in Figure 3. For computational vision systems, replicating this kind of perception was a significant obstacle for an extended period. It was not until 2019 that Convolutional Neural Networks (CNNs) started to effectively model phenomena akin to Gestalt's principles<sup>4</sup>. At the time of this writing (early 2023), advanced image analysis systems, including those included in language models like ChatGPT, are more and more able to employ these principles of perceptual organization.



Figure 3. This time ChatGPT correctly recognizes this image as a shoe.

The principles of human perceptual organization explain many well-known visual illusions as well, like the Mueller-Lyer illusion, in which two identical segments are perceived as having different lengths, or the Beau-Lotto color illusion, in which identical patches of color are perceived as having different hues. However, calling them illusions does not give full justice to the fact that they are the manifestation of human natural processes in vision responsible for perceptual constancy, detection of depth, distinction between object's color and environment's luminance and so on. If we see the world *as it is*, as Ruskin advocated, we would not be subject to such perceptual illusions. But the crucial point is the fact that those biases are an integral part of the way we perceive the world and, consequently, they are also part of the way we aesthetically experience it. This means that systems programmed to analyze and generate content that are then experienced and enjoyed by humans must also integrate models of how humans perceive.

To move to another example, consider these two versions of the Mona Lisa face: one has been modified to give the face a sadder expression.<sup>5</sup>



Figure 4. Sad and happy Mona Lisa.

Humans can read subtle changes in facial expressions. In (A) Mona Lisa's countenance appears sadder than the original version (B): the facial expression in image A distinctly varies from B *overall*; this difference is particularly pronounced in the mouth and eyes, which convey a dissimilar appearance. But if you cover the mouth in both images, you will see that all features of the face are actually identical, and that the sole alteration lies in the mouth's corner<sup>6</sup>. Essentially, we almost "hallucinate" an illusory expressive change that does not exist. This phenomenon occurs because our top-down interpretation of an emotional state (happy vs sad) subtly alters our overall perception of identical shapes. Today, artificial visual systems can classify (or generate) human expressions and attribute affective labels like "sad" or "happy". However, such systems may not perceive the (illusory) overall change in expression as we do in this instance. A skilled artist, by contrast, understands that even a minor alteration in a feature, such as the curvature of the mouth's corners, can transform the entire image's perceptual impact. This understanding arises from the artist's own perceptual experience, which mostly aligns with that of the audience. Similarly, for an artificial system to be effective, it must incorporate a model of human perception, including its peculiarities and distortions. Such integration is essential for the system to understand and generate content that successfully captures these subtle sensory effects.

Not only top-down cognitive knowledge can tinge our perception, but affective and expressive features greatly determine how we see or hear something. For instance, a melody in a minor key may sound "sad", a bright color may seem "cheery", and a jagged line might convey "nervousness". These qualities, while not physically inherent

in the stimuli, are consistently perceived across individuals and cultures. Moreover, they are very difficult to separate from our overall perceptual experience: I cannot hear the melody “innocently”, that is, taking out its sadness; I cannot look at the jagged line without perceiving at the same time its nervousness. Gestalt psychologists refer to these as “look-and-feel” or “expressive” qualities (*Anmutungsqualitäten*). The application of these qualities in understanding aesthetics and art is pivotal<sup>7</sup>.

Now, the expectation for AI to replicate human subjective “feels” might be misplaced. Currently, debating whether artificial systems can experience perceptual qualities as humans do is not particularly fruitful. This would suggest the need to attribute subjective experiences, private states of consciousness, and individual perspectives to machines. The philosophical discussion surrounding “qualia”—the unique sensations of personal experience—is already sufficiently complex within the context of human consciousness; expanding it to include artificial systems isn’t yet necessary. What is more pertinent is exploring the ability of artificial systems to model and predict human experiential processes, especially those related to aesthetics, art, and emotions. For example, machine learning already enables the association of shapes with their affective qualities, training systems to categorize, for instance, “jagged” lines as “nervousness” or minor-key music as “sadness”. Affective image classification has been advanced in recent years. Algorithms can be trained with images from sources like the International Affective Picture System (IAPS), which are pre-evaluated for emotional impact, to classify the emotional attributes of new images. Today, if we ask Midjourney to generate a “sad” landscape, the system will accordingly generate some image depicting a scene that exudes sadness. If we ask GPT to analyze a joyful image, the system will add in its description that the scene and the colors are joyful, and so on. Affective computing, similarly, focuses on creating systems that can recognize, interpret, and simulate human emotions, using technologies like facial recognition algorithms, voice analysis, and other biometric sensors. In art history, the classification of affective body postures is used to analyze the distribution of every *pathosformel* and its prevalence within an artistic tradition, facilitating computational methods in exploring affective archetypes in Western iconography, a study rooted in Warburgian tradition<sup>8</sup>.

## Perceptual expectations: the historicity of the eye

As we just saw, affectivity, feel and mood tinge the way we see the world. And knowledge and cultural background not only help us classify what we see, but enrich and change the impression of what we see<sup>9</sup>. This leads to the concept of the historicity

of perception: the idea that individual and cultural experiences shape our view of reality. Different periods and backgrounds lead to varying interpretations and aesthetic perceptions of cultural artifacts. Styles, as art historian Michael Baxandall called it, reflect a “period eye”. Culture and experience shape how the brain processes visual information, determining what is aesthetically significant and influencing how artists represent the world. Art becomes a key witness to how an era perceived reality and analyzing an era’s visual culture allows us to reconstruct its worldview, where style changes signify not just an evolution of artifacts’ formal properties but also shifts in collective perception.

Consider how contemporary reactions to designs from different eras, such as 1970s interior design and fashion, vary. Today, these styles are often labeled as “tacky”, “old-fashioned”, or “heavy”, and are seldom considered “fresh”, “modern”, or “elegant”. However, an observer from the 1970s might strongly disagree with these characterizations. Despite being materially identical, these designs evoke different moods and expressive qualities over time. They are perceived differently by someone from the 1970s compared to an individual of today’s era.

Perceptual differences extend beyond the temporal distance to include also cultural differences. Two individuals from different artistic backgrounds will react differently - emotionally but also perceptually - to the same piece of art. This leads to what was mentioned in Chapter 1 regarding the importance of precise and individualized customization of taste and aesthetic preferences through algorithmic analysis of audience behavior. This customization acknowledges that each subject perceives and reacts to content uniquely, with variations that are clustered around temporal, spatial, and cultural differences.

Another aspect of the historicity of perception is our ability to trace the stylistic evolution of objects. Take, for example, laptops or iPhones from different years. When we compare them, we do more than just note their similarities. Viewing an object as a variation or an evolution of its predecessor alters our perceptual feeling about it: We tend to see newer models as contemporary evolutions of their older counterparts. Our understanding of design history informs our aesthetic judgments of these changes. This dynamic is similar to how we perceive a modern rendition of an old song: we listen to the new version through the lens of the original. However, this perception shifts if we are unfamiliar with the original. For instance, children who first become acquainted with a cover version of an old song often perceive the original as a variation of what they already know.

Historicity and temporal sequentiality (and the different kind of perception they elicit) add a layer of complexity of how artificial systems might model what humans “see”. There are promising developments in analyzing large art datasets to address artistic influence in music<sup>10</sup> and painting<sup>11</sup>. By identifying patterns and considering the chronological order of artworks, we can trace the origins and influences of novel patterns. Given machine learning’s potential, these systems might even surpass humans in uncovering patterns of similarity and historical influences. By analyzing human aesthetic responses, they should also integrate how style and form evolution influence the observer’s perception and emotional reaction. These systems must account for the diverse affective nuances and interpretations unique to each observer.

## Calibrating to Human Imperfection

While we do not expect that an artificial system sees and “thinks” like a human being, given its role in studying and generating content intended for human interaction, it must have nevertheless a sense of how humans perceive and react. This touches on what is defined in Chapter 8 as the problem of “AI alignment” in the context of aesthetics. This also requires an understanding of the *limits* and *biases* natural to human perception and cognition. Artificial systems can analyze data beyond human capabilities, but it remains crucial for them to be able to take into account the typical modalities of human perception.

For example, when an AI system trained in object recognition fails to discern details in an image, we might conclude that either the system lacks sophistication, or the image is excessively blurred or noisy. However, in the realm of human aesthetic experience, moments of confusion are not necessarily flaws to be fixed. Often, an object captivates our interest precisely because it poses a perceptual or cognitive challenge, such as ambiguity or indeterminacy. Eric Kandel writes: “Just as the artist creates a work of art, so the viewer recreates it by responding to its inherent ambiguity. The extent of the beholder’s contribution depends upon the degree of ambiguity in the work of art”<sup>12</sup>. Artistic techniques like “estrangement” and defamiliarization (*ostranenie*) thrive on uncertainty and interpretative instability, sparking the viewer’s curiosity and attention. **Unlike in AI, where ambiguity might be a problem to solve, in aesthetics, ambiguity is a feature.** An artwork’s aesthetic richness often lies in its resistance to a singular interpretation, remaining open to multiple readings.

Consider for example George Braque's painting "Violin and Pitcher" (1910), analyzed by Michael Baxandall in 1994 (Figure 5)<sup>13</sup>. This Cubist painting blends figurative elements with abstract ones, particularly in what Baxandall refers to as "the mysterious left flank". This section of the painting, difficult to process, creates a perceptual tension where the eye oscillates between recognizable elements (the violin, the pitcher, the nail) and the more enigmatic parts. Baxandall terms this a "perceptual lock", valuing the painting for its ability to challenge and stimulate the viewer's perception, creating a "cognitive itch"<sup>14</sup>.



Figure 5. Appreciating the confusion.

In this context, indeterminacy arises both from the painting's inherent complexity and the human observer's cognitive limitations in discerning shapes and features. Our capacity for pattern recognition has its bounds, and we often find ourselves lingering in the uncertainty of an artwork, accepting and embracing this indeterminacy as part of the aesthetic experience. This poses an interesting dilemma for AI. Machines excel in labeling and pattern recognition, optimizing detection in ways that might surpass human capabilities. However, during an aesthetic experience, this kind of optimization

may not always be desirable or relevant to understanding the human way to perceive and react.

---

## **Unlike in AI, where ambiguity might be a problem to solve, in aesthetics, ambiguity is a feature**

In conclusion, to model human aesthetic perception effectively, AI systems need to accomplish two tasks. First, they should recognize configurations that may evoke a sense of ambiguity in human observers. Second, when generating images or other media content, they should strategically employ ambiguity or uncertainty to enrich the viewer's aesthetic experience.

### **Aberrant pattern recognition and “data unconscious”**

Imagine again a machine learning system tasked with object recognition that occasionally misidentifies what it sees. As said, these systems, used in computer science for categorization, detection, and prediction, aim for high precision. Annual competitions assess which system most accurately identifies objects in photographs or videos. Potential errors in identification are not only signs of imperfection; they also raise concerns about biases that may replicate societal stereotypes and cultural assumptions, leading to adverse social impacts.

In contrast, within the human domain of aesthetic experience, *aberrant* interpretation of information has long been acknowledged as a catalyst for creativity. Thus, research in AI aesthetics should also consider atypical classifications not as mere errors but as potential creative insights. 2015 Google's Deep Dream, where images are reinterpreted through layers of unexpected shapes, was reminiscent of avant-garde movements that embraced free association as tool of enhancing artistic innovation, like Surrealists and Dadaists. Techniques such as automatic writing, dream as a source of inspiration, and random processes enabled them to transcend the limitations of conventional categorical structures shaped by cultural and experiential norms, thereby fostering serendipitous creativity.

Similarly, when a machine learning system misclassifies an object, it might inadvertently uncover an association between two entities. This association, grounded in an underlying similarity or connection, may be previously unnoticed and, while potentially incorrect, is not senseless. Consider this example (Figure 6), where the image of the Queen of England is analyzed by a rudimentary neural network<sup>15</sup>. The misclassification by the network of the Queen's crown as a "shower cap" reflects the infancy of machine learning algorithms in grasping object recognition. While at first glance this error may elicit amusement, it also subtly reveals a creative analogy. The crown and the shower cap, while diametrically different in value and function, share a visual similitude that the algorithm has inadvertently highlighted. Moreover, as we will briefly discuss, this mistake might have an impact on our perception and interpretation of the picture itself: at a closer look, that crown actually looks like a shower cap after all.

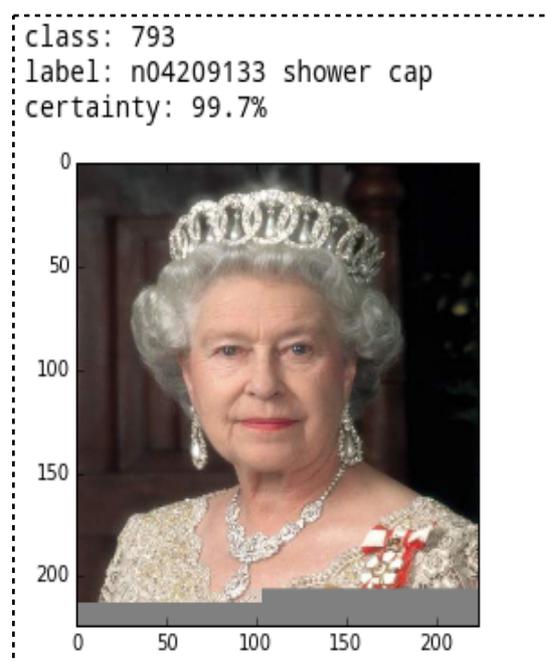


Figure 6. A fancy shower cap.



Figure 7. Artist Georgia O'Keeffe (1887-1986) "holding her new cell-phone".

Another notable instance of an imperfect image recognition system is the one interpreting Alfred Stieglitz's 1918 photograph of Georgia O'Keeffe as "a man holding a cellphone" (Figure 7)<sup>16</sup>. This blatant anachronism, where contemporary technology is perceived in an old photo, reflects our modern perceptual habits ingrained in the image recognition system. Here again, even those familiar with the photograph may find themselves unable to "unsee" this new interpretation. Although erroneous, these descriptions do have an influence on our ways of viewing these images. Associative mechanisms have always been a crucial factor in the flux of cultural and historical reinterpretations.

While computers are trained for accuracy and objective reality representation, in the realm of aesthetics, AI can facilitate the discovery of subtle, perhaps historically unfounded, associations, unveiling intriguing connections. Notable examples are projects like "MosAlc"<sup>17</sup> and "X Degrees of Separation"<sup>18</sup>. The first one involves a sophisticated AI algorithm created by researchers at MIT, designed to identify parallels between artworks originating from vastly divergent art historical periods. This algorithm conducts comparative analyses of pieces from the Rijksmuseum and the Metropolitan Museum of Art. The second example is a collaborative artistic venture between Google and artist Mario Klingemann. This innovative online application employs artificial intelligence to construct a seamless transition between two images, ensuring that the intermediate images represent a coherent and natural progression from the start to the end image.

Particularly, unsupervised neural networks can sift through vast data sets without the confines of pre-established guidelines. This freedom enables the emergence of alternative classification systems that are sometimes more advanced or intricate than those devised by humans. In science, for example, these systems can bring order to complex phenomena, such as meteorological patterns in cloud formations, by adopting innovative classification schemes that surpass traditional methods.<sup>19</sup> In the realm of cultural analysis, platforms like Spotify exemplify this approach. They analyze musical tastes, trends, and styles, transcending the limited range of familiar musical genres through algorithmic data processing. This technology can differentiate thousands of genres, enhancing the granularity of classifications. Categories can become not just more precise but also fluid and continuous, reflecting a shift from traditional and discrete classifications to a more dynamic and expansive framework.<sup>20</sup>

**While contemporary discussions rightly focus on the risks of such systems in extracting, reinforcing, and perpetuating human stereotypes and biases present in their training data, it's also worth considering their imaginative and exploratory potential.** These systems can discover new patterns and offer fresh perspectives. The question arises: Can an AI create a novel sensibility, and if so, can we as humans perceive and understand it? Walter Benjamin introduced the concept of the “optical unconscious”, suggesting that media technologies such as photography and cinema reveal layers of reality not ordinarily accessible to the naked eye, much like psychoanalysis uncovers hidden impulses. These mediums, through techniques like magnification, montage, slow motion or detail isolation, offer new perspectives on perceptual reality we otherwise had not access to. Big data analysis and machine learning advancements herald a similar potential. These technologies can discover unseen patterns and connections in complex cultural phenomena, both historical and current. AI data analysis is a powerful tool to unearth a “data unconscious”, uncovering patterns that often elude our perception due to human perceptual and cognitive limitations. This raises also the question: Are these patterns merely overlooked aspects of our reality, are they comparable to a kind of “archaeological” excavation of our cultural material? Or, alternatively, is this process akin to the imposition of patterns, like the way constellations were perceived and *constructed* from randomly scattered stars by various cultures?

Connotations and free associations in aesthetic experiences carry a range of implicit meanings and emotional resonances, shaped by our cultural and personal backgrounds. Free association, the spontaneous connection of thoughts and emotions, often reveals unexpected links. When AI systems misclassify or generate “strange patterns”, they might be tapping into this same associative impulse, uncovering

relationships that evade human comprehension. These algorithmic “slips” challenge us to consider whether those associations are just random or if they emanate from a deeper, yet-to-be-understood order.

---

**While discussions rightly focus on the risks of AI systems in reinforcing human stereotypes and biases present in their training data, it's also worth considering their imaginative and exploratory potential**

### **Artificial Platonism and counterfactual imagination**

One central feature of intelligence, whether natural or artificial, is the ability to abstract. Advancements in artificial intelligence lead to the emergence of systems capable of learning from images and songs, extracting and interpreting the *essence* of styles, artifacts or motifs. These AI models, fueled by deep learning techniques, analyze vast collections of visual or auditory artworks, discerning patterns and stylistic nuances inherent in different genres or epochs. In the project described in Chapter 1, the characteristics of hundreds of faces in Rembrandt paintings were analyzed and extracted, and then used to generate a new image in Rembrandt's style. Similarly, by training a network with thousands of Bach's chorales, a neural network was able to produce its own Bach-like chorales. In all these examples it should be noted that the system does not generate just an average of the paintings (or of the musical works) that have been analyzed, nor does it copy parts of the various Rembrandts by recombining them like a collage. Instead, it extrapolates the general patterns of Rembrandt's style, including his brushwork and his use of colors. What is reconstructed and reused in a generative key is the *quintessence* of the painter's work, the Platonic idea of a Rembrandt. The newly generated painting is one materialized instance coming from the *latent space* of all possibilities determined by this essence.

A consequence of the development of those “quintessence machines” is the transformation of artifacts defined by their uniqueness into instances of a general idea, that could be potentially materialized in an infinity of variations. We could see this as an advanced kind of technical reproduction, in the vein of Walter Benjamin's famous

essay. However, it differs fundamentally from the concept of reproduction as merely copying a specific individual work. Instead, it involves producing new variants by reproducing the same style, motifs, “vibes”<sup>21</sup> of an author’s body of work. The essence of the author’s creative signature is distilled and then used as a template for creating new, original pieces that echo the original style yet diverge in their individual expression.

There are also legal implications, as only individual works are protected by copyright, not an artist’s style. The potentially infinite reproductions in the style of Rembrandt, Bach or Beatles by machine learning systems do not infringe copyrights, allowing to generate new melodies in their style that everybody is free to use. This is the reason why commercial platforms already offer the possibility to upload songs and ask a machine learning system to generate new melodies that mimic the style of the input music, making possible to avoid any possible copyright issues<sup>22</sup>.

The tradition of rip-offs, *hommages* or outright imitations in the cultural industry is age-old: for many producers it is easier to capitalize the trend of some successful musical, narrative or cinematographic innovation than create something new. However, one might wonder if the increasing ease in extracting the quintessence of works by means of AI-generative systems won’t raise the need to protect not only individual content, but also styles, aesthetic “feeling” or moods, or even abstract ideas underlying an artist’s oeuvre.

The relationship between an object and its “essence”, between an object and its abstract and generalized idea, has relevant perceptual and cognitive implications. Consider this imaginary scenario of a world devoid of knowledge of geometry, a place where the principles of shapes and forms are yet to be discovered. In this hypothetical world, a prominent artist leaves behind a legacy of just eight abstract paintings. Each masterpiece displays configurations that, to our real-world eyes but not to the eyes of this imaginary world, resemble with what we know as “triangles” (much like Malevich’s renowned “Black Squares”). In this world, these shapes have no geometric interpretation and are perceived merely as eight unique and somewhat similar elements, characteristic of the artist’s recurring motifs and style. Enter an advanced algorithm, similar to contemporary AI systems, that meticulously examines these paintings. The system manages to extract their underlying pattern and formulates an abstract concept: the geometric idea of a triangle. This allows the system to generate an endless array of triangles-paintings, all mirroring the artist’s distinctive style. Suppose the eight painted triangles are all equilateral or isosceles; the artificial system now

introduces us to the possibility to generate scalene triangles. Intriguingly, some might find these AI-generated variations more appealing than the “originals”.

This imagined scenario brings us back to Plato’s idea that reality is only an imperfect reflection of perfect ideal entities. Each real triangle depicted in those eight paintings is only an imperfect manifestation of the abstract concept of triangularity. But the crucial point is that we now look at the artist’s images differently: before, we used to see those triangles as unique objects in their singularity. Now, instead, we consider them as singular occurrences among numerous potential arrangements. By extracting the essence of those objects, we now see the actual singular images as examples of the vast array of possibilities within the latent space defined by the machine - each image a path taken from many not pursued.

*Artificial Platonism* arises from the belief that artifacts are entities that embody a specific essence that can be extracted and used. This concept becomes particularly relevant when such methods are applied to complete unfinished works or extrapolate missing data. For instance, when an AI system attempted to complete Schubert’s Symphony No. 8, the “Unfinished”, it generated numerous possible continuations. These variations were derived from the latent space determined by the essence of Schubert’s musical style. Ultimately, programmers selected the variation they deem most fitting. To assess the AI’s effectiveness in completing a piece of music, one could also provide the AI with only a section of an otherwise complete work and then have it generate the remaining part. The AI’s output can then be compared with the actual remaining portion of the piece (a sort of statistical cross-validation). But an AI-system could be also trained not only to generate a “Schubert simulation”, but also to optimize its output for higher aesthetic and musical appeal and be an “enhanced Schubert”.

**The use of AI in culture suggests that each cultural product is just one possibility within a vast latent space of alternatives, some of which may be equally or even more interesting or good.** Consequently, we might argue that while Franz Schubert made specific artistic choices in his compositions, other paths were equally viable within the scope of his stylistic essence. We might call this *counterfactual imagination*. Not only does this perspective allow us to view an object as one among infinite possibilities of an abstract idea, but it also compels us to contemplate the different forms that object or image could have taken. As we engage with art and artifacts through this lens, we are not only acknowledging their existing state but also actively imagining the myriad ways they *might have been realized differently*. This shift in perception underscores the fluidity and multiplicity inherent in cultural creations, the boundaries between the actual and the possible become increasingly blurred.

## **The use of AI in culture suggests that each cultural product is just one possibility within a vast latent space of alternatives, some of which may be equally or even more interesting or good**

### **Strange hands: a digression**

Staying in our imaginary scenario of a world lacking geometry, we assumed that the eight triangles depicted in the works of art are only isosceles or equilateral. However, the artificial system, in order to create variation, applies the concept of triangle in a more general way and thus generates scalene ones, with all different sides. This production of never-seen-before triangles would lead to two scenarios. The first, which has already been discussed, is that the system opens up a new world of possibilities for observers, possibilities that were not considered before and that have been made possible by the creation of a general model of a triangle. The second, which we briefly discuss here, is that, for the observer of that world, the artificial system just *makes a mistake* because it generates triangles that seem nonsensical to them, lacking the regularity they were used to. This would reveal how, in that world, the acceptable level of abstraction of triangular objects (defined by the eight paintings) was not arbitrary but limited to thinking of them only as isosceles or equilateral, with no possibility of expanding to new triangular shapes.

This situation mirrors the early phase of AI-image generation when systems like Midjourney repeatedly made errors in drawing hands (particularly getting the number of fingers not right), putting too many teeth in a human mouth or generating images of texts in bizarre fonts and non-existing writing systems. These weren't specific flaws in the systems' ability to draw hands or texts; they generate these elements in the same manner as they draw trees, clouds, and rocks. *The difference lies in our perceptual expectations*: we have more precise expectations for hands and texts than we have for trees, rocks or clouds, due to their well-defined characteristics in the real world, such as the typical five fingers on a hand or the structured nature of written characters within a known alphabet and language.

Image generation systems, in their training phase, encounter a variety of hand representations in images. Often, some fingers are obscured or hidden, while in instances of clasped hands, more than five fingers may be visible. Consequently, these systems learn to perceive the hand not as a fixed entity with precisely five fingers, but rather as a flexible collection of fingers, in the same way as how they interpret a tree, seeing it as a flexible collection of branches and leaves. As a result, the systems generate images of hands with a number of fingers that can vary, reflecting this learned abstraction rather than the precise anatomical reality of human hands. Similarly, the representation of written texts in image generation systems may lack the linguistic and alphabetic precision we expect, resulting in sequences of characters that seem random or non-standard. The challenges encountered by these systems consists therefore in capturing and reproducing the specificity and precision of our expectations for these objects, in contrast with elements like trees, clouds or rocks, for which we accept greater variability and generality in representation. If we viewed hands and written text with the same degree of variability as trees or clouds, their unusual representation might appear normal to us, as it might to an external observer like an alien not yet accustomed to the specifics of the human body and writing. These peculiarities reveal more about the features of our perception than about anomalies in image generation systems.

## **Déjà vu and the *sensorium's* shifts**

As we have observed, AI-generated content can seem unusual or even uncanny, primarily because it defies our perceptual norms. The uncanny arises when we encounter anomalies in otherwise familiar contexts – for instance, hands with six fingers, biomorphic shapes melding into objects, or facial and bodily expressions that are slightly off. However, there are instances where we might expect novelty or unconventionality, yet what we encounter feels eerily familiar, akin to a *déjà vu*. This phenomenon is often noticeable in artificially created artworks. To those who are not experts, distinguishing between an original and an AI-produced work can be challenging, leading to a sensation of vague familiarity.

Let us examine the images depicted in Figure 8, produced by a Generative Adversarial Network (GAN) trained on a dataset of late 19th-century paintings. It is noteworthy that, despite being informed of the images' artificial origins, some observers not only perceive a typical modern painting in the Impressionist style, reminiscent of Renoir's portraits, but also experience a sense of *déjà vu*, as if they have previously encountered

these paintings. The sense of familiarity stimulated by the GAN painting could be explained by the fact that the image is indeed composed of iconographic and stylistic elements of images that we have probably actually seen in the past. If those images were generated according to platonic abstraction of style and motifs, *déjà vu* here is analogous to Platonic *anamnesis*, according to which, all experience and knowledge are a remembrance of essences that have always existed (see Chapter 3).

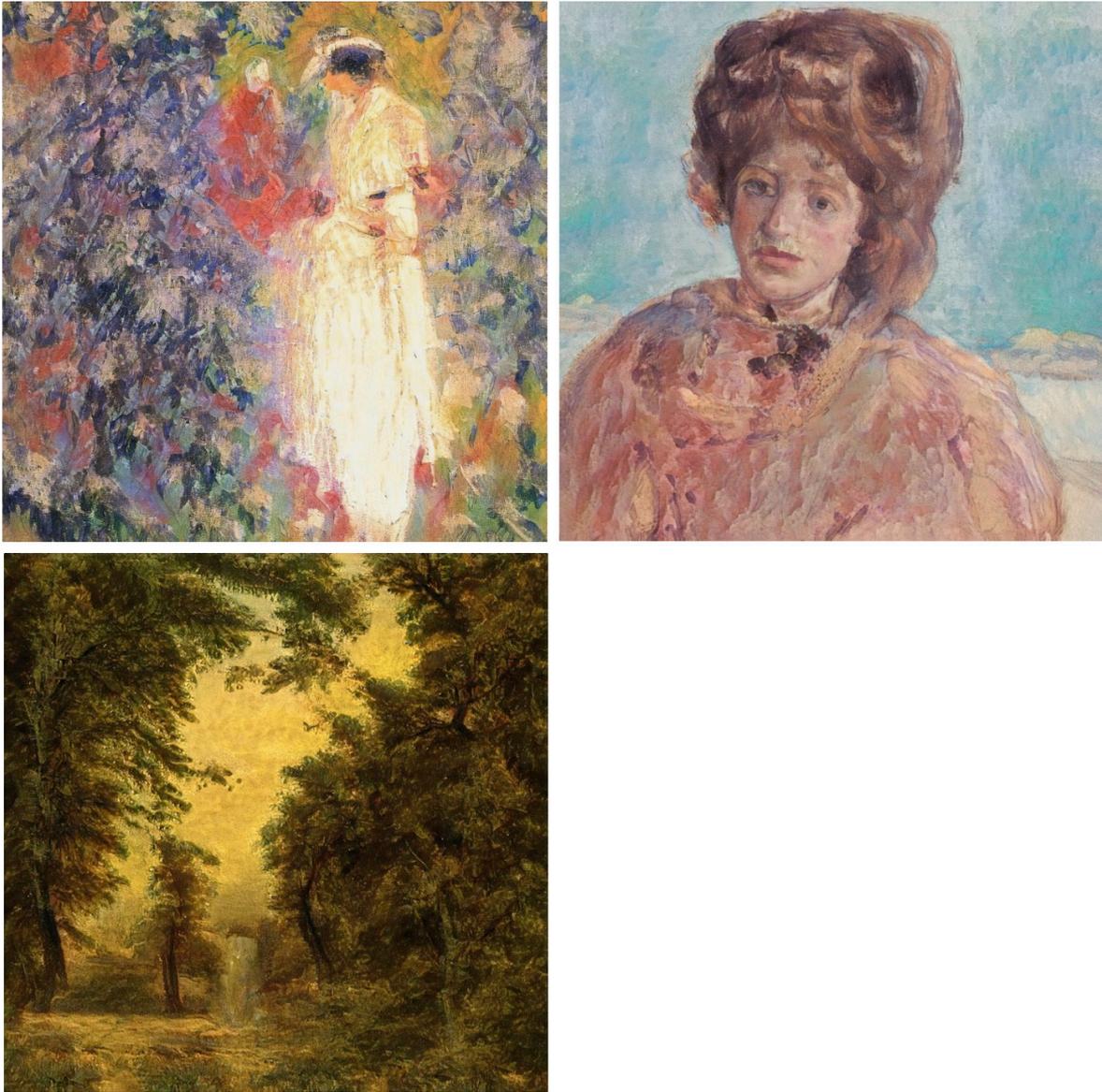


Figure 8. Images in Impressionist style generated by GAN in 2021.

In the perception of familiarity, the relationship between original contents and their artificial variants to which subjects are exposed plays a crucial role. In contexts where we primarily engage with original cultural products, artificial creations are often seen

as mere curiosities or derivatives. However, it's conceivable that in the future, content generated from original works may surpass them in abundance.

This shift in the relationship between original content and (artificial) variations can be delineated in two general phases:

1) new content generated is just viewed as a reflection or extension of the original works. For instance, when we see an image as "in the style of Picasso". To refer to our previous hypothetical scenario: at first, in each newly generated artificial triangle, we recognize in it the eight famous triangles we are familiar with. *The variations are perceived through the lens of the original objects (we see the original in the variations).*

2) Following an intermediate phase, where it becomes unclear whether what we observe is a derivation of original works or original in itself (creating a déjà-vu effect), we may progress to a stage where artificially generated content predominates over the original. Continuing our example, upon encountering one of the eight triangles painted by the renowned artist, an individual may perceive it simply as another instance of the general concept of a triangle, indistinguishable from others previously seen. In this case, *the original is viewed and identified through the prism of its various iterations (we see the variations in the original work).*

This phenomenon bears a striking resemblance to musical *remixes* and *cover versions*, as previously discussed. Some reinterpretations gain popularity by re-introducing old songs to a younger audience unfamiliar with the originals. Older listeners discern the original within the cover version (often with a nostalgic preference for it), while younger listeners, upon encountering the original, may in contrast perceive it through the lens of the cover (often preferring the latter, as it aligns more closely with contemporary tastes and styles).

The remix example shows that recreating original works or their styles is, of course, not a new concept. For instance, there was a long debate about which of Rembrandt's self-portraits were genuine and which were just imitations by his students. This debate was only recently settled, with the number of recognized originals reduced from 90 to 40. Similar trends of imitation can be seen in music, whether it's Baroque or 1960s Pop, where certain styles and sounds become popular and are imitated by artists. What's different now is how easily and frequently current AI systems can create works "in the style of" famous artists and the shifts that can occur if images, sounds and text become mainly the product of artificial generation. By creating endless variations and blending

of the products of human culture, we might get to perceive the original works and content as just instances of an endless array of potential artifacts. In our imaginary world without geometry, the original eight triangles now appear as being just manifestation among the others of the general idea of triangle. Similarly, the 40 Rembrandt's portraits, 400 Bach's chorales, or about 200 Beatles' songs, once seen as unique and special may be viewed as just instances of many variations of a certain general style.

Moreover, AI-generative technology facilitates forms of blending, recombination, and recontextualization of styles and forms, making it easy to generate images whose style is, for example, a mixture of Hieronymus Bosch and Andy Warhol, or a blend of Caravaggio and Manet. Similarly, one could have a Beatles song follow melodic evolutions that incorporate Bach, or a Shakespearean sonnet speak of the drama of a financial collapse, and so on. The very possibility of freely generating these variations has an impact on our way of seeing the cultural objects we are accustomed to. We are now better at naturally "interpolating" between styles and contents with our own eyes thanks to the training we are subjected to. These new possibilities also change how people who use or know about these technologies see art. In other words, **AI technologies generate novel patterns and classifications that we partially integrate into our thinking and perception** making it easier for us to envision, even before asking the system to generate such images, how a painting by Hieronymus Bosch might manifest in a Pop-Art or postmodern composition (Figure 9) or to conceive of a contemporary desk in the style of 17th-century Flemish art (Figure 10). In this sense, **AI systems train us in new ways of seeing.**



Figure 9. Dall-E 3 image generated with the prompt: "Blend Hieronymus Bosch with Pop-Art."



Figure 10. Dall-E 3 image generated with the prompt: “A PC-set in the style of XVIIth century Dutch-still life.”

This phenomenon mirrors the historical influence of innovations in media devices, which have consistently reshaped human cognitive experiences. Walter Benjamin explored how technologies alter the human *sensorium* by introducing new ways of processing and engaging with the world<sup>23</sup>, particularly by diminishing the “aura” of a work of art, its unique presence in time and space and its authentic history. With the advent of technologies capable of mass reproducing art, these works can now be experienced in different contexts and locations, thereby changing the original perception that was tied to the artwork’s uniqueness and authenticity. Later Marshall McLuhan notably extended this view: each medium, with its specific characteristics, affects our sensory balance. Electronic media, particularly television, fostered a more holistic and less linear perception of the world compared to the linearity of print media. The experience of artificial generation of content, where boundaries between shapes and styles are made much more fluid, enhances this transformation. Moreover, McLuhan argued that the content of any medium is always another medium; it’s the characteristics of the medium itself that shape and alter the sensorium, changing not just what we see, but how we see by altering the balance among our senses. Similarly, the true extent of the changes that “generative AI” can induce are still in their nascent phase and will be the task of future investigations.

**AI technologies generate novel patterns and classifications that we partially integrate into our thinking and perception.. In this sense, AI systems train us in new ways of seeing**



Figure 11. Closeup photo of a bee.

## Notes

<sup>1</sup> John Ruskin (1857), *The Elements of Drawing*.

<sup>2</sup> Image from Lisa Feldman Barrett (2016), *How Emotions are Made: The Secret Life of the Brain*, Houghton Mifflin Harcourt.

<sup>3</sup> As the Deep Dream developer and Google engineer Alexander Mordvintsev, explains: “Intuitively, this means changing the image rather than changing the network [...] so that the resulting image is shaped by what the network ‘expects’ to see”. Further: „A defining feature of the Deep Dream algorithm is the use of backpropagation to alter the input image in order to minimize categorization errors. This process bears intuitive similarities to the influence of perceptual predictions within predictive processing accounts of perception,” <https://www.nature.com/articles/s41598-017-16316-2>.

<sup>4</sup> See <https://link.springer.com/article/10.1007/s42113-021-00100-7> ; [http://bengio.abracadoudou.com/cv/publications/pdf/kim\\_2019\\_arxiv.pdf](http://bengio.abracadoudou.com/cv/publications/pdf/kim_2019_arxiv.pdf). See also: Kriegeskorte, N. (2015), Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Song, Lichtenberg, & Xiao, 2015; Bai, 2011; Girshick, Donahue, Darrell, & Malik, 2014; Ren, He, Girshick, & Sun, 2015), <http://www.cs.toronto.edu/~sven/Papers/CCN2019a.pdf>

<sup>5</sup> Based on O. Schwartz, H. Bayer, and D. Pelli (1998), “Features, frequencies, and facial expressions”, *Investigative Ophthalmology and Visual Science* 39, 173.

<sup>6</sup> This works better if the two images are viewed separately and not next to the other.

<sup>7</sup> As shown by Rudolf Arnheim in the seminal book *Art and Visual Perception*, University of California Press, 1954.

<sup>8</sup> L. Impett and F. Moretti (2017), “Totentanz : Operationalizing Aby Warburg’s ‘Pathosformeln’”, *Stanford Literary Lab: Pamphlets*, 16, <https://litlab.stanford.edu/LiteraryLabPamphlet16.pdf>

<sup>9</sup> D. Stokes (2014), “Cognitive Penetration and the Perception of Art”, *Dialectica* 68: 1–34. D. Stokes (2018), “Rich Perceptual Content and Aesthetic Properties”, *Evaluative Perception*, edited by A. Bergqvist and R. Cowan (Oxford: Oxford University Press), 19–41.

<sup>10</sup> Giovanna Castellano, Eufemia Lella, Gennaro Vessio, <https://arxiv.org/pdf/2003.08476.pdf>; <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-019-0214-8>.

<sup>11</sup> <https://arxiv.org/abs/1408.3218>.

<sup>12</sup> Eric R. Kandel (2012), *The age of insight: The quest to understand the unconscious in art, mind, and brain, from Vienna 1900 to the present* (New York, Random House), p. 192.

<sup>13</sup> M. Baxandall, ‘Fixation and Distraction: The Nail in Braque’s *Violin and Pitcher* (1910)’, in *Sight and Insight: Essays on Art and Culture in Honour of E. H. Gombrich* at 85, ed. J. Onians, London, 1994, 401. Referenced in John Onians, *Neuroarthistory*, Yale Univ Press, 2007.

<sup>14</sup> Baxandall, 'Fixation and Distraction", p. 413.

<sup>15</sup> <https://codewords.recurse.com/issues/five/why-do-neural-networks-think-a-panda-is-a-vulture>.

<sup>16</sup> "Dream Formulations and Deep Neural Networks: Humanistic Themes in the Iconology of the Machine-Learned Image," kunsttexte.de, 2017, <https://edoc.hu-berlin.de/bitstream/handle/18452/19403/Spratt%20-%20final.pdf>.

<sup>17</sup> <https://www.csail.mit.edu/news/algorithm-finds-hidden-connections-between-paintings-met>.

<sup>18</sup> <https://artsexperiments.withgoogle.com/xdegrees/>.

<sup>19</sup> As noted by Kurihana et al. (2022), "Unsupervised learning enables us to move beyond artificial categories derived from historical cloud classification patterns, facilitating the discovery of more nuanced classifications." In "Cloud Classification with Unsupervised Deep Learning", <https://arxiv.org/abs/2209.15585>.

<sup>20</sup> See Every Noise at Once, [everynoise.com](https://everynoise.com).

<sup>21</sup> To borrow from Grietzer's "Theory of Vibe," <https://www.glass-bead.org/article/a-theory-of-vibe/>.

<sup>22</sup> See, for instance, [aiva.ai](https://aiva.ai).

<sup>23</sup> "Thus technology has subjected the human sensorium to a complex kind of training", in "On Some Motifs in Baudelaire" (1939).