Emanuele Arielli

Chapter 9. Made *by* and *for* humans? The issue of aesthetic *alignment*

Chapter 9 of *Artificial Aesthetics: Generative AI, Art and Visual Media* by Lev Manovich and Emanuele Arielli. Published October 27, 2024. 9,527 words. All book chapters: http://manovich.net/index.php/projects/artificial-aesthetics-book

> "AI is, in large measure, philosophy. It is often directly concerned with instantly recognizable philosophical questions: What is mind? What is meaning? What is reasoning and rationality? What are the necessary conditions for the recognition of objects in perception? How are decisions made and justified?"
> Daniel Dennett[1]

"I am not a robot" and the problem of demarcation

Artists and artisans are terms derived from *artifex*, which refers to someone who creates something "artificial" as opposed to what is natural and not made by humans. For example, a natural item could be a coconut shell used to collect water and from which we drink. In contrast, a cup or a glass would be an artisanal product, and today we use them instead of coconut shells. However, not all advancements replace what came before. Industrial mechanization led to the mass production of objects that were previously handcrafted, such as dishes, chairs, furniture, and clothing. But artisanal production continued, albeit in different forms, as niche creations for those who appreciate human effort and prefer it over mechanical seriality, or simply because they enjoy the process of crafting artifacts. Similarly, the invention of photography did not lead to the disappearance of painting. However, a new medium often brings about significant transformations of the old one. Painting remained relevant in the post-

photography era because it moved beyond naturalistic realism and differentiated itself from photography through innovative styles and concepts.

Consider the difference between an industrially produced item, like an Ikea vase, and a handmade one. We often value handmade and artisanal products more highly, attributing to them a superior quality compared to industrial goods, even if this is not always guaranteed. Beyond quality, we perceive these items as having an 'aura' of uniqueness, with physical characteristics that cannot be replicated. While it is possible to produce industrial items with features that mimic handmade qualities, such as imperfections and slight variations in shape, we generally have a strong aversion to 'faux artisanal' items because they are seen as faking an old mode of production. Furthermore, as discussed earlier in the context of effort, we project onto the handcrafted object the commitment of its maker, viewing the object as an embodiment of the time spent and the technique learned over the years. The object thus becomes a carrier and witness of a human presence that is no longer evident in industrial items.

Similarly, the advent of AI doesn't mean the end of non-AI creations. However, it could transform how we view and use them, affecting their economic value and distribution. An artist who doesn't use AI might initially feel obsolete compared to one who does. Yet, the role of non-AI craft could be redefined as a specific market emerges for "artisanal" works valued for being human-made, much like the appreciation for "handmade" objects developed following industrial serialization. We might see a renewed interest in what machines cannot replicate, encouraging works that highlight human imperfection and uniqueness. This could lead to a new appreciation for traditional craftsmanship and art forms emphasizing the individual artist's touch, contrasting with the perfection often associated with AI-based creations.

Old techniques survive thanks to their ability to differentiate themselves from new ones. Just as painting distinguished itself from photography through experimentation beyond realism, artisanal production is characterized by not appearing serial and perfect like industrial production. In the context of generative AI, it's unclear whether we have a new medium capable of generating distinct content and how previous media might differentiate themselves. This explains the attention given to cases where AI-generated content is nearly indistinguishable from non-AI content, particularly in image generation.

While a classic Turing test evaluates whether machine-produced content can pass as human-made, the challenge of *demarcation* involves developing criteria to distinguish

AI-generated content from human-made content when potential indistinction is the norm. This includes creating criteria that ensure human-made content is genuinely human (an “inverse Turing test”) and that AI-made content is indeed produced by AI (an “AI-originality test”).

In aesthetic production, several well-known cases challenged the intuitive ability to distinguish between AI-generated and non-AI products:

a) Consider the case where an AI-generated image was used in a competition that did not specify which digital tools were allowed. Jason Allen won first prize in the digital arts category at the Colorado State Fair Fine Arts Competition in 2022 with a work titled “Théâtre D'opéra Spatial”, created using Midjourney.

b) Another example involves an AI-generated photograph presented as real. At the 2023 Sony World Photography Awards, Boris Eldagsen won in the “Creative” category with “Pseudomnesia: The Electrician”, a black-and-white image. After winning, Eldagsen revealed the image was indeed AI-generated and refused the award to raise awareness about the impact of AI technologies in the artistic context[2].

c) There is also the case of a real photo that was presented as AI-generated, winning in a category reserved for AI-generated images at the Creative Resource Collective (CRC) Photography Awards in 2023. Photographer Miles Astray won third place in the AI category with his work “F L A M I N G O N E”. However, it was later revealed by the artist that the image was actually captured with a traditional camera and was not AI-generated. As a result, the CRC withdrew the award[3].

d) Another case involves a real photo by Australian photographer, Suzi Dougherty, that was disqualified from a local photo contest after her submission was mistakenly thought to be AI-generated. The contest, organized by Charing Cross Photo, aimed to highlight local fashion photography. The organizer of the contest explained that he disqualified the photo due to its perceived artificiality, noting that the mannequins and Dougherty's shot appeared “too perfect” [4].

The situation where human productions must prove their authenticity, ensuring they are not artificially generated, is comparable to the situation in which websites require users to prove they are human. A CAPTCHA (“Completely Automated Public Turing test to tell Computers and Humans Apart”) is a type of challenge-response test used in computing to distinguish between humans and automated programs, aiming to prevent

bot attacks and spam. Although these systems were initially designed to pose questions that only humans could answer, artificial systems can now pass these tests without significant limitations. In principle, it is getting easy to artificially simulate human behavior[5], while it is quite difficult for a human to simulate artificial behavior (e.g., passing a speed calculation test or generating a truly random sequence[6]). The utility of CAPTCHA still lies in the fact that humans are slower, which is enough to slow down high-frequency, high-intensity artificial attacks. In fact, a person who responds to a CAPTCHA too quickly might raise suspicion. Similarly, in Suzi Dougherty's case, the too perfect and polished nature of the photo and the models' poses raised suspicions that it was machine-generated. In a competition where only AI-produced images are allowed, an "AI-originality test" should be able to exclude content that is actually human-made. This is however challenging since AI-generation can mimic human content as well.

A comparison with chess helps clarify the issue. In chess, a player likely hasn't used a computer if their moves show errors and imperfections. On the other hand, very sophisticated and "perfect" moves might suggest computer use[7]. In this case, an "inverse" Turing test checks if someone is human by looking for "imperfections", but this is a limited approach since a machine can deliberately play less perfectly to mimic humans. Moreover, this approach could become harder also because humans, who train with computers, could adopt more computer-like playing styles.

Content type	**Appears AI-made**	**Appears human-made**
AI-made content	Passing the "AI-originality" test[8]	Passing the Turing test (or: deceiving the "inverse" Turing test)
Human-made content	Deceiving the "AI-originality" test	Passing the "inverse" Turing test

If we examine the case of text generation, it is quite challenging to determine if a text was written using a Large Language Model (LLM). In the early phases of their diffusion, it became quite common to consider the presence in a text of verbose expressions like "delving into the intricate tapestry etc." as an indication that an LLM was involved. AI-based applications exist, however, that can automatically "humanize" the text, as well as AI platforms that help recognize if a text is AI-written. This situation leads to an interesting consequence: if someone wants their text to be perceived as LLM-generated

(and pass an "AI-originality" test), they might simply put these stereotypical formulations in their texts. On the other hand, writers who naturally use such expressions might feel the pressure to avoid them in order to ensure that their work appears human-made, potentially changing their writing style for good: today, one might avoid terms like "delve" or "intricate tapestry" in their writings to prevent raising suspicion of AI intervention. This would be an interesting case of the impact of these technologies on human writing style. Similarly, the diffusion of AI technologies could also lead in the domain of visual aesthetics to a conscious avoidance of certain styles when someone wants to highlight the human, non-technological aspect of their creations. For instance, images that are too smooth and perfect, like Dougherty's fashion photograph, might be avoided to prevent the false impression that they were created with AI, even when this is not the case. "Artisanal style" in craft also involves keeping the irregularities and unpolished features well visible to reduce the suspicion of machine assistance.

When it becomes difficult to distinguish between content produced with or without AI, more advanced methods are needed. These often involve machine assistance, much like how machine learning is used to identify forgeries and artistic replicas[9]. As with the Voigt-Kampff test in the science fiction movie *Blade Runner*, where androids were identified by means of detailed questions triggering affective reactions, telling apart human and AI-generated language will become harder as LLMs improve. We already rely on programs that can judge whether a text was probably written by AI or a human, detecting subtle features that might go unnoticed by the reader. Similarly, it is possible that we will need artificial tools (or a combination of human expertise and AI systems) to determine whether visual works, musical compositions, architectural designs, or television series were produced with significant use of generative AI.

The case of Astray's "Flamingone" photograph is a paradigmatic example of non-AI-generated work misrepresented as AI-generated. This belongs to a type of deception – presenting as artificial something that is not - that goes back to pre-technological eras: a prime example is the "Mechanical Turk", created by Wolfgang von Kempelen in 1770. This device appeared to be a chess-playing machine but actually concealed a small-sized human chess player inside. It amazed audiences because it seemed to be a mechanical system playing chess at a high level. The trick wasn't to make something artificial appear human, but rather to use a hidden human to make a machine seem extraordinarily capable.

Today, a company could market content as AI-generated even when it is not: behind the fascination of the label "powered with AI", they might rely on more traditional technological systems, or worse, employ human assistants – like Kempelen's hidden chess player - who actually perform the tasks. In a scenario where the specific difference between works made with or without AI technologies can no longer be detected, we might have to rely solely on forms of guarantees and certifications that something was made by human (or, conversely, with AI-help), and thus refer to the historical process that produced that content (its "provenance" or historical traceability, according to Jaron Lanier[10]). This is similar to the problem posed by deepfake photographs and audiovisuals: when they become indistinguishable to both the human eye and deepfake detection models, only the traceability of an authentic origin will serve as the imperfect criterion of demarcation.

However, we can also envision a scenario where the issue of demarcation becomes entirely obsolete. This could happen for two main reasons. Firstly, the integration between individuals and technology, which already exists for traditional technologies, may come to include AI as part of the normal processes of human production. For example, we might see practices where AI's formal suggestions become inspirations for artisans such as woodworkers or ceramic craftspeople in the physical creation of their works[11]. Secondly, the impossibility of true demarcation may lead to a "post-artificial" situation, as discussed in the previous chapter, where we ultimately suspend judgment on the true authorial origin of a work, permanently abandoning the question of whether something is genuinely "made by humans" or not.

An image is worth 60 words: language as a paintbrush

Approximately 4% of the population is estimated to be "aphantasic," meaning they are unable to have visual mental images (or auditory imagination) [12]. These individuals think and remember in an abstract and verbal way only. Some may have partial mental imagery or experience visual information in dreams, while others lack visual mental imagery altogether. Aphantasic individuals often describe what they saw using language and factual knowledge rather than visual recall. They may provide detailed accounts based on their understanding of the object or scene rather than a mental image of it. This condition highlights a spectrum of visual imagination capabilities in humans. On the opposite end of this spectrum, in fact, we have people who can vividly visualize an image and therefore describe it verbally in a precise manner[13].

This neuroscientific phenomenon illustrates how the relationship between language and images can vary individually, but also how language can have different functions in its relationships with images: in some cases, language is used to merely describe an already present image, while in others, it has the function of *generating* the image, for example in the listener's imagination.

In generative AI, this leads us to the specific case of text-to-image (TTI) interfaces (but also text-to-music, or text-to-text), where a person uses verbal descriptions to suggest to the AI system what to generate. This is a process that involves the generation of high-quality images by means of numerous iterations of verbal prompts, which is almost an art form in itself and a specialized means of communication between users and AI. While prompts can be quite specific, according to some sources, they generally do not need to exceed the limit of about 60 words[14].

This process is similar to the traditional work of a forensic artist who sketches a person based on a witness's verbal descriptions: the forensic artist carefully listens to every detail the witness provides, as the witness tries to recall an image and translate those fragments of memory into words. It is both a *reconstructive* and an *interpretative* work: the witness's words describe an image they are attempting to recall but also reconstruct something that needs further definition. The sketches created by the artist are not just direct translations of the witness's words; they also reflect the artist's stylistic choices and interpretations, which reshape the witness's mental image. This creates a continuous loop in which the witness's descriptions shape the image the artist draws, which in turn influences the witness's memory or vision, blurring the lines between creation and recollection.

Closer to the aesthetic dimension is the communication between a client and a designer: for example, a person commissioning an architect to design a house or interior decor. In these cases too, it's naive to think that the client already has a clear idea of what they want and is simply helped by the professional designer to clarify their vision. The relationship between client and designer in this sense is not purely instrumental: the designer tries to satisfy the client's needs, mostly expressed in words, but often has the authority in aesthetics and design skills to provide new ideas and guidance, influencing the client's choices.

Words can take on different roles in relation to the image: they are used to describe an image we already have in mind, but they can also be tools to make the machine imagine for us. To clarify this point, we can refer to two classic rhetorical figures concerning the

relationship between text and image: *ekphrasis* and *hypotyposis*. Ekphrasis (literally "description") consists of using language to describe an image, specifically an artwork. This is what you get when you ask an AI system (like ChatGPT) to analyze a picture, but also what we do when we want to convey to a TTI system an "image" that we have in mind in a clear and definitive way. Hypotyposis, on the other hand, means to "outline" or to "draft," which stresses the aspect of "generation," as trying to induce with words an image. Hypotyposis, traditionally, involves creating vivid mental imagery through language, aiming to evoke strong sensory and emotional experiences in the reader. In essence, ekphrasis is about describing an existing image, while hypotyposis is about letting generate an image through description.

The ekphrasis perspective suggests that the user has a clear image in mind and uses text-to-image systems as a tool to realize this internal vision. The user crafts a detailed description to guide the machine toward producing the specific image they envision, making adjustments and refinements as needed. Conversely, the hypotyposis approach starts with the user providing a vivid linguistic description without a specific image in mind, relying on the system to generate an image based on this description, effectively "imagining" it. In this process, the machine significantly influences the final visual output, as the user has not yet formed a clear vision of the desired outcome.

This distinction is theoretically significant because casual users often treat text-to-image systems as simple generators of verbal ideas, whereas professionals use them to describe something precise they have in mind. During the iterative cycle of generation and refinement, what was initially unclear can become increasingly defined, thus shifting the approach from simply "letting the machine generate" to "describing to the machine" what it should produce. This represents a progression from hypotyposis to ekphrasis.

In converting text into visual images, a significant issue in TTI systems is the limitations of language and cultural differences in visual communication. The categorical constraints of language can limit these models by restricting the range of concepts they can accurately represent. For example, if a language lacks specific words for certain colors or shapes, the resulting images might not fully capture the intended details or abstract ideas. Therefore, the limitations imposed by language affect the model's ability to create images that align with human creativity and intent, revealing a gap between textual descriptions and visual output: "whereof one cannot speak, thereof one cannot generate the imagery," to paraphrase a classic sentence by Ludwig Wittgenstein.

Not everything that can be imagined—even vaguely—can be accurately translated into words. In fact, there may be forms of imagination, inspiration, or moods that lack linguistic equivalents. Periodic and iterative refinement allows us to overcome these limitations, bringing us closer to the idea we have in mind but cannot precisely describe. Moreover, using words entails significant constraints both individually (as people have different capacities and styles of expression) and culturally (as different languages have semantics that do not align and describe similar concepts differently). An emblematic case is the visual interpretation of complex emotional states. The interpretations will inevitably be influenced by the AI's training dataset, which might favor specific cultural associations. Given that the expressive, aesthetic, and emotional lexicon varies from language to language, when verbal description is used, it can lead to outcomes where the categorical differences of language end up consolidating in the production of images as well.

Obviously, the categorical limits of language should not make us lose sight of the fact that interfaces between users and generative AI can also function without linguistic aid. Although this type of interface has been dominant from 2022 to 2024, it is only one of many possible modes of interface and input in the application of so-called diffusion models and transformers. The possibilities for “instructing” the system range from choosing from pre-configured palettes of styles, aesthetics, and “vibes,” to inserting images from which to draw inspiration in terms of composition, light, or style, to using sketches and drafts.

On conceptual AI-art

In its early stages, AI focused on forms, images, and the sensory dimension of objects. On the level of forms, AI's potential lies in its ability to extract, manipulate, and combine patterns, whether in images or music. But *ideas are also patterns*, specifically structures of concepts, mostly codified through language, and just as AI can link or merge similar visual patterns, it can also easily manipulate, combine, or identify similarities in conceptual structures. The shift to systems capable of processing, reformulating, and creating text allows AI to work not only with forms and sensory objects but also with ideas, concepts, and discourses. In this sense, AI art expands into its conceptual and symbolic dimension. Today's large language models demonstrate unprecedented abilities in natural language processing, reasoning, and creative tasks. These models can engage in debates, generate complex narratives, and even suggest “new ideas”.

Language models are already able to explore existing information and cultural content and to suggest interesting conceptual connections, also in artistic domains.

Aesthetics is understood as the domain where sensual impact plays the most relevant role in determining our judgment of something being agreeable, beautiful, striking, powerful, astonishing and so on: however, in last century's avant-garde, artists definitively broke with the ideal of sensorial beauty, considering it rather a matter of superficial decoration and an obstacle for freedom of expression. The idea of something "beautifully crafted" fell under suspicion: art overcame the necessity of the artist's craft. Ready-mades and other re-appropriation of everyday objects (as in Duchamp or, later, in Warhol) made evident that there is no perceptual feature that distinguish an artwork from common entities: it is not necessary for art to be sensorially striking, but it need to be meaningful, to be *about* some symbolic idea or concept being embodied by the artwork itself.[15] For example, Walker Evan's photographs (1936) might appear physically identical to Sherry Levine's *After Walker Evans* (1981) appropriation of Evan's photographs. However, they express different ideas and therefore they are different artworks.[16] If ideas are crucial for an artwork, then aesthetics should not limit itself to be a theory of sensorial appearances, but needs also to develop to an *aesthetics of meanings*. This means that not only good or beautiful appearance, but originally structured and relevant ideas as well, conveyed through interesting and engaging forms, are necessary for having great art. Conceptual art, despite its focus on ideas, still relies on an expressive form—otherwise, simply stating ideas would be enough. Traditional art, on the other hand, was never purely about aesthetics; it always involved a conceptual layer. Without this, art would be reduced to mere decoration, limited to pleasing images of portraits and landscapes. What has shifted over time is the balance between form and concept, with the conceptual aspect gaining prominence in contemporary art. This evolution has placed a stronger emphasis on the originality and novelty of ideas as key measures of artistic value.

In his manifesto on conceptual art, Sol Lewitt (1968) famously wrote: "The idea becomes a machine that makes the art". On a similar vein, machines could produce ideas useful for art: an "artificial conceptual artist" would be a system that is able to find relevant and meaningful links between concepts and ideas and find out original ways to put those links into a sensible form. An "idea generating"-machine would even be proficient in generating so-called "art-speak" discourse. Language models can produce text that could synthetize arguments and narratives that a critic, a curator and the artist usually produce to convey meaning to cultural products and artworks. AI would be able to generate conceptually rich art-discourse, helping individuals to develop consistent

curatorial/critical assertions that can enhance the interpretation and the aesthetic judgment of artworks. At the same time, AI's role as a "collaborator" might extend to conceptual discovery, identifying non-obvious connections between ideas. Those systems can act as a "sparring partner" and a brainstorming tool, allowing artists and thinkers to test and refine their ideas in a dynamic exchange.

Prompt: Think of a conceptual performative art piece

"Transcendence through Erasure"
This conceptual performance piece explores the idea of artistic transcendence through the radical act of self-erasure.
The performance begins with the artist silently entering the space, carrying a large bucket of black paint. Without any preamble, the artist proceeds to systematically cover every surface in the performance area - the walls, the floor, even the ceiling - in thick, opaque black paint.
As the space is gradually consumed by the encroaching darkness, the audience is left to grapple with the implications of this action. Is the artist destroying the space, or transforming it into something new? Is this an act of creation or destruction?
Once the entire space has been blanketed in black, the artist turns their attention to their own body. Dipping their hands into the remaining paint, they begin to methodically cover themselves, until their form is indistinguishable from the surrounding environment.
The audience is now confronted with a void - a space devoid of visual stimuli, where the artist's physical presence has been subsumed by the darkness. The only indication of the artist's continued existence is the occasional shift in the shadows or the faint sound of their movements.
As the performance continues, the audience is invited to explore the transformed space, guided only by their other senses - the smell of fresh paint, the slight vibrations in the air, the muffled echoes of their own footsteps. The familiar becomes unfamiliar, and the audience is forced to re-evaluate their understanding of the performance and their relationship to the artist.
Gradually, the artist may begin to emerge from the darkness, revealing their painted form in a slow, almost ritualistic manner. This reappearance, however, is not a return to the initial state, but rather a symbolic rebirth - a transcendence through the act of erasure.
(generated by Claude 3.5 Sonnet)

Machine judgment: Beauty is in the AI of the beholder

As mentioned in the first chapter, AI systems have a dual capability of creation and *evaluation* of content. AI technologies have shown remarkable proficiency in identifying patterns, recognizing their subtle variations, and discerning differences that often elude human perception. This ability extends not only to analysis and classification but potentially to aesthetic evaluation, where machines could theoretically formulate judgments and critical analyses that might surpass those of the average person. Even by simply inputting an image into a large language model like ChatGPT and asking it to evaluate the image's qualities, the system can provide a thorough iconological and visual analysis, list possible historical and artistic references, and suggest its strengths and weaknesses. AI might potentially generate sophisticated interpretations of artworks, and this could change the role of human art critics.

In the field of design, machine learning is increasingly used to optimize design choices across a wide range of complex constraints, generating variations from which creators can draw inspiration. In complex creative contexts such as architecture, the integration of AI-systems into design processes is leading to a new approach where the machine's analytical capabilities assist professionals in testing ideas while ensuring compliance with constraints, such as physical requirements, statics, legal regulations, and environmental standards. Software can explore numerous alternatives, optimizing for factors like material efficiency and structural integrity. This trend suggests a future in which *machine judgment* will play an increasingly important role in design decisions.

AI systems can evaluate the aesthetic quality of visual content by learning from human judgment. Therefore, these systems can also *predict* how people, given an image, would rank content in aesthetic value.[17] One example is a tool like Everypixel's neural network, which assesses the aesthetic value of stock images. This system assigns scores based on visual quality and optimizes search results by prioritizing higher-scoring images. This approach is not novel; as early as 2017, AI Mirror utilized Google's Neural Image Assessment (NIMA), a convolutional neural network trained to predict the aesthetic appeal of images. The NIMA model was trained on large datasets like the AVA dataset, which contains over 255,000 images rated by amateur photographers.[18]

AI-systems can develop a sense of what is aesthetically pleasing and artistically relevant not only by making use of rankings by users, but by directly accessing the corpus of texts *speaking about* art or design: the vast amount of text data may be transferable to aesthetic evaluation. An AI-model can even adjust the metrics of aesthetic evaluation to

specific prompts given by the user in order to assess images according to the user's specific taste and preferences. [19] Moreover, while individuals often have a sense of their aesthetic preferences, they may struggle to articulate the specific reasons behind their choices. AI systems could offer insights into these preferences by analyzing user's observed choices and then elaborate a model of the user's aesthetic taste and suggest more refined aesthetic judgments.

However, the challenge for AI lies to use human aesthetic criteria, using individual judgments as learning benchmarks. This raises the question about which aesthetic criteria those systems need to be trained in. In AI research, the concept of "ground truth" denotes the reference data used as a benchmark for evaluating the performance of an algorithm or model. It represents the "reality" that the AI system is trying to model or predict. For example, in the context of image recognition, ground truth could be the precise label of objects in images, annotated by humans. But what are the ground truths of aesthetics? The simple answer would be: specific human responses to stimuli, like appreciation, affective and cognitive judgment, liking in front of specific artifacts. Ideally, those responses should allow for building a model of human aesthetic sensitivity and be able to predict how humans would react to new forms or artifacts (see Chapter I). Since aesthetic judgment is also dependent on general cultural values, symbols and traditions, those models should also theoretically be able to describe and predict human psychological and cultural sensibility. One difficulty is the fact that aesthetic preference and taste, as well as critical judgment concerning artworks, show great inter- and intra-individual variations based on personal experience and historical context. Aesthetic universals seem to be confined to very general perceptual qualities, but aesthetic preferences seem to vary in space and time, and to change also for a person in different moments of her life. For instance, a particular stimulus may be perceived as pleasant initially, but its appeal may diminish as it becomes too predictable. Similarly, someone might initially prefer decorative elements in design but later develop a taste for minimalism, viewing the same decorations as overly ornate or sensorily overwhelming.

More interestingly, aesthetic evaluation in people is also negatively defined by our perception of what *bad taste* is. Judgment of bad taste, or *Kitsch*, is partly determined by social factors, class membership and the implicit desire to set us apart to those we consider culturally and social different. Some artifacts could be perceptually pleasant, like a photograph of a sunset on the ocean, a panoramic urban skyline, or a black/white wallpaper of a sleeping baby, but we may judge those images as too stereotypical. Mostly, these images might be judged as clichéd or intellectually unengaging, suitable only as prefabricated wall art rather than as objects of deeper aesthetic appreciation.

Machine learning systems, as mentioned, build their model using data from user's judgment in online photography platforms. Those judgments could greatly diverge in assessing what is beautiful and what is *Kitsch*, and sometimes the same image could be judged in both ways. An artificial system having a sense of what could be considered "bad taste" should consequently be able to differentiate *for whom* an artifact appears to be *Kitsch*, taking also external factors such as cultural context and social distinction as determinants of such judgments. There is no one single "ground truth" reference in aesthetic judgment, and social factors in human aesthetic appreciation should be integrated in models of artificial evaluation and generation of artifacts.

This also shows the limitation of artifact generation that is based on the extraction of *average* aesthetic preferences: While a sunset might generally be considered more beautiful than a trashcan, or a high-contrast photo better than a shaky one, relying solely on these average criteria risks producing artificial kitsch. A too easy and standardized taste could be appreciated at the beginning but becoming dull later; true innovation and interesting art often emerge when average expectations are violated.

The issue of aesthetic alignment

One crucial consideration in this regard is the fact that AI systems, like humans, operate on internal models of the world that may not perfectly align with reality. The common assertion that AI "sometimes" hallucinates is, in fact, an understatement. AI systems are continuously generating outputs based on their trained models, and these outputs can be considered a form of constant "hallucination". The key is that these hallucinations often correspond closely enough to reality or human expectations to be useful or convincing. The same can be said of humans, who also operate with models of the world that are imperfect and prone to errors. Our ability to interact effectively with our environment is largely due to the fine-tuning of our perceptual and cognitive systems over millions of years of evolution. Similarly, AI systems must be fine-tuned to human aesthetic sensitivity, but this process is far from straightforward.

To program AI models that can attune to human aesthetic preferences, these systems must be calibrated to reflect the dynamic nature of human experience. This calibration must account for individual and cultural variations, as well as the fluidity of tastes and trends. A machine that models a human evaluator by learning from human judgment belongs to the domain we called "subject generations" (Chapter I).

A critical distinction emerges when we consider the sources of information that shape AI and human internal models of the world. While AI systems are typically trained on vast digital datasets, human perception and cognition are the products of millennia of evolutionary adaptation to the physical world. The brain of an infant is created with a DNA that guides the development of its nervous structures while it is in the mother's womb. The ancestors of that individual going millions of years in the past interacted with the world through sensory perception and evolutionary selection allowed the genes that encode the more useful perception and classification process to be passed on.

On the other side, the internet, on which AI-systems are trained, provides an incomplete and often misleading representation of reality. Even if the entirety of the web would be used as training data base, the real world is orders of magnitude more complex. Humans, too, when susceptible to forming their understanding of the world based on internet-derived information, may not always align with the complexities of real-world experience.

As philosopher Merleau-Ponty (1908-1961) emphasized, human perception and cognition and our understanding of the world are fundamentally shaped by our physical and sensorial interactions with it. This perspective suggests that to truly align AI with human aesthetic sensitivity, we may need to move beyond purely digital training and incorporate *embodied* experiences. According to this perspective, perception should not be considered just passive reception of data but an active engagement with the world, shaped by the body's interaction with its environment. For AI to align with human aesthetic sensitivity, it would need to be similarly integrated into the world, potentially through the development of AI systems embedded in physical bodies. These embodied AI systems would interact with the world in a manner akin to humans, having not just sensory experiences but also having to deal with the human ecosystem of norms, conventions, and social dynamics.

This could also mean envisioning artificial agents endowed with drive and motivation, with integrated aesthetic goals and preferences. While this might seem far-fetched, we could speculate how this immersion could potentially lead to the development of a sense of agency and authorship within these AI entities. In the aesthetic context, this would require a functional definition of what it means for a system to strive for aesthetic pleasure and having edonic preferences, possibly rooted in the recognition of harmonious forms or mechanisms linked to aesthetically pleasing patterns, the satisfaction of predictive mechanisms[20], or the achievement of an optimal balance between uncertainty and familiarity.[21]

Such systems would not merely respond to stimuli but would actively seek to model the world in ways that reflect goals and desires. However, the question arises if AI should be limited to merely replicating human preferences and aesthetic sensitivity. We could argue that AI could potentially go beyond human aesthetics, creating entirely new forms of beauty that humans might not have conceived but that could fulfill machine goals and preferences. These machine-generated aesthetics could involve levels of complexity, that are not designed for human consumption.

Consider an AI system generating music. If it were aligned with average human preferences, it would produce melodies that mimic simple structures, such as the verse-chorus-bridge format commonly found in pop music. On the other hand, if the AI were to truly push beyond human preferences, it would generate compositions that had an extreme harmonic complexity, akin to some sort of avant-garde composers like Arnold Schoenberg, who developed the twelve-tone technique to further scramble harmonic expectations and are object of aesthetic consumption and appreciation for only a minority of well-trained people. Theoretically innovative music could just turn out to be unbearable to listen for people who are used to more traditional harmonic structures. The AI's alignment would need to balance innovation with accessibility, potentially including settings to adjust the complexity of musical structures according to a listener's taste or providing the public with opportunities to develop a corresponding appreciation for these aesthetic innovations. Alternatively, when translated into a form accessible to humans these creations might be "dumbed down" and simplified to meet our perceptual and cognitive capabilities, in the same way that a chess program might have a setting that allows it to self-limit, enabling it to play in a manner that is manageable for the human player.

To make a further example, a typical novel might follow a linear plot with clear symbolic references and an AI-system which is too aligned with general human preferences might produce work that feels formulaic. However, AI might produce a text that layers multiple narratives, each with its own set of symbols and meanings, akin to James Joyce's *Finnegans Wake*. The AI might generate dense, intertextual references that only advanced readers or literary scholars could fully appreciate. On the other hand, if it exceeds human cognitive capacities, the literature it generates could be too esoteric, requiring a dumbed down version for broader accessibility.

Synthetic data and "AI cannibalism"

The scenarios just described, although fascinating from a speculative point of view, seem to be moving in the opposite direction to what many observers and critics appear to see in current trends. Rather than a "collaborator" with surprising capabilities assisting the artist, many foreshadow a dispossession of the creative class towards a generic homogenization of content. Rather than systems capable of opening alternative paths, AI systems would just learn from mediocre databases and tend towards a generic kitsch in his generations, stemming from the average of human preferences. To avoid a scenario in which AI runs out of high-quality data, training databases need to be carefully curated and developers need to implement strategies that continuously incorporate fresh, diverse, human-created content into training datasets. However, according to critical views, if AI-generated content endangers the careers of artists this could lead consequently to a diminished influx of new, diverse artistic styles, essential for training and improving of AI models.

Studies have suggested that, paradoxically, while generative AI may enhance individual creativity, it could also reduce the collective diversity of novel content. This phenomenon is partly due to our increasing reliance on AI systems that offer templates and pre-packaged solutions.[22] This critique emphasizes that in the current use of generative AI, especially those systems based on prompts, the user does not work — as Michelangelo famously described — by starting with a rough block of stone and "removing" the unnecessary material to reveal the ideal form the artist has in mind. Instead, users start with a default setting of images that often features kitsch element, like idyllic landscapes, rich and colorful palettes, or stereotypical magazine cover models that depict humans. To create works of genuine aesthetic value, the user must begin with these iconographic stereotypes—products of widely accepted standards of beauty—and must possess the capacity to achieve sufficient "escape velocity" to break away from kitsch and impose their own vision.

Moreover, the pervasive use of AI in selecting and distributing visual content may even have a feedback effect on our development of taste and preferences in the direction of uniformity. For some commentators, we are already seeing a visual aesthetic convergence, for example in the style of cafés and hotel design and in the curated looks of Instagram-inspired interiors. From this perspective, the homogenization of aesthetic experiences, amplified by AI-content, could potentially stifle creativity and diversity in visual culture.[23]

To address the issue of systems that rely too much on average preferences extracted from the training data set, researchers try to distinguish between *general* aesthetic assessment - the average or most common aesthetic preferences in a given domain - and *personalized* aesthetic assessment.[24] The former refers to the analysis and modeling of average or widespread aesthetic preferences in a specific domain of interest, particularly images. The latter focuses on the analysis of data from an individual towards modeling their specific aesthetic preferences. This enables AI systems to predict a person's rating of new content and even generate content tailored to their aesthetic preferences in a similar fashion as how algorithmic recommender system in video, music or consumer product do.

The trend in personalization, on the other hand, may be also troubling. With AI systems tailored to ensure that a given profile has a completely unique content, we could move toward to a stage of *hyper-personalization*. This could result in aesthetic *echo chambers* where users would be fed with content that would just satisfy their own tastes and preferences, thus further limiting them from exposure to other aesthetics. Aesthetic products as collective and shared culture-building phenomena would be compromised by excessive individualization of the content each person would consume by means of tailored AI-generation.

Finally, another trend that is source of concern is the potential for AI-generated content to feed back into training datasets, creating a self-referential loop. Future data training sets will increasingly consist of AI-generated outputs, such as blog posts, articles, images and even fiction, as these types of content become more and more prevalent on the internet. According to recent research, this recursive loop could be disastrous for the stability of the models.[25] For example, AI-generated paintings, which are already trained from the history of human art, might then be included in future training datasets. The next generation of AI, trained on this mixed dataset of human and AI-generated art, would produce works that are even further removed from original human creativity. This self-referential loop has been referred to as "AI cannibalism," and the gradual degradation over time has been dubbed "Habsburg AI" (a reference to the Habsburg dynasty, known for the recurrent inbreeding of its members), point to the fact that an "inbreeding" of data could result in a gradual loss of diversity, originality, and quality in the generated content.[26]

The risk is amplified by the growing use of *synthetic data* in machine learning training. Synthetic data refers to artificially generated information used to train AI models when real-world data is scarce, expensive, or difficult to obtain. The use of synthetic data is

particularly prevalent in fields where real-world data is limited or sensitive, such as medical imaging or rare event simulation. While this approach is often necessary and beneficial in these contexts, applying similar methods to creative fields risks homogenizing the aesthetic landscape.

The self-referential "cannibalism" of systems that learn from their own outputs and rely on synthetic data raises the risk of *qualitative* degeneration, but it could also have negative effects just from a *quantitative* perspective. We can recall the case of composer David Cope (Chapter 8), who, to overcome a creative block, began developing a system in the 1980s that could generate thousands of musical compositions in a specific style. Generative AI enormously facilitates the transition point between an imagined idea, vaguely present in our minds, and its realization. However, this ease in content generation can lead to hyper-production and an inflationary flood of content that does not necessarily translate to creative abundance or innovation. Instead, it can result in potential perceptual and cognitive overload.

The richness of creative possibilities can risk getting lost in an ocean of options. As for august 2023, more than 15 billion images were created using text-to-image algorithms. Stock photo companies are gradually adding AI-generated images to their catalogs. This is the quantity of photographs that were produced in 150 years, from the first photograph taken in 1826 until 1975.[27] From this perspective, AI and bot-generated content, mainly text and images, could soon surpass human-generated content simply because they are easier to produce. Considering that writing is a time-consuming activity for people, a similar fate could occur for textual production, whether literary, journalistic, or academic. Machine-generated content is continuously growing and is also becoming the basis for training future language models.[28]

This happens, moreover, in a scenario where, even without AI, produced content in some domains already surpass demand for consumption. Academic production of papers is one specific example. In popular culture, for instance music, SoundCloud as for mid-2024 has 350 million tracks by 40 million artists, and, according from data from 2023, more than 120.000 tracks were uploaded to streaming services *every day*.[29] The question therefore arises if we do really need AI songs, if human produced music may have well reached a saturation point, and so many other fields of creative expression is drowning in over-abundance.

Aesthetics for machines

People may continue to create images, texts, songs, and books for the personal satisfaction of the act itself, but the question is whether this abundance of content will capture enough human interest. In a context of limited attention and time to evaluate cultural over-production (be it images, songs, or books), even before the advent of generative AI, what consequently might emerge could be the need for "superhuman" capabilities in analyzing and evaluating human content. The use of AI for evaluation and judgment, therefore, would serve both to allow AI to learn from the data provided to train it and to compensate for people's cognitive and temporal limitations in analyzing the overproduction of content. As AI amplifies our ability to produce content while simultaneously exacerbating the problem of absorbing that content, we might then delegate back to AI the task of understanding and elaborating on that content.

In everyday textual communication we can already observe a similar loop: as AI language models facilitate and amplify the production of texts, articles, emails, and messages, those tasked with reading, evaluating, and responding to this influx of information may increasingly rely on AI to synthesize, process, and, if necessary, respond to these communications. This loop risks reducing us to mere facilitators in a dialogue between machines.

While this might seem like a pessimistic view of the future, it underscores a critical point: AI systems still depend heavily on high-quality, human-generated content for effective learning and development. The traditional relationship between humans and tools has been largely one-sided, with humans benefiting from their use: machines, including AI-systems, are extensions and interfaces of human activities (Chapter 8). However, as machines increasingly take up human-like tasks, this dynamic is shifting. Artificial systems benefit from human input, with humans *acting as interfaces* or extensions of these systems in the world.[30] Human behavior and content help extend and refine the capabilities of artificial systems, becoming the source of training for such systems. This topic has been the subject of extensive discussion and controversy, particularly concerning the legitimacy of exploiting human labor as "fuel" and raw material to power systems that ultimately aim to eliminate the need for that very labor.

One aspect of this development is that AI systems not only produce aesthetic artifacts for humans but also influence humans to create content with machines in mind. As machines increasingly evaluate aesthetic value and rank content, there is a growing tendency to adapt to machine judgment. This is already evident when musicians

compose tracks that align with the preferences of streaming platforms or when content creators tailor their work to appeal to algorithms. Some may view the rise of non-human evaluators as encouraging the creation of artifacts and content optimized according to criteria established by machines. As these systems become more prevalent in evaluation, suggestion, value ranking, there is concern that they could prioritize machine-defined standards over human creativity. However, these systems might actually enhance our autonomy by encouraging us to break free from established patterns: creativity has always been subject to constraints—cultural habits, technical and material limitations, dominant trends, social conformism, and the need to meet others' expectations all influence human thinking. Artificial evaluative systems could be thoughtfully designed to address these decision traps in creative thinking.

Moreover, we should not overlook the possibility that we might *prefer* being judged by a machine over a human, in the same way that we are less hesitant to submit incorrect and very rough text drafts to a large language model compared to a human reader. In some contexts, like therapy, studies seem to show that people are less self-conscious and more willing to open up to therapeutic chatbots than to human therapists, suggesting that interacting with a machine reduces inhibitions because there is no fear of personal judgment.[31] Similarly, creating for a machine - rather than immediately exposing one's work to human evaluation - might make us feel freer and more willing to experiment. In fashion, for instance, we might be more inclined to experiment with outfits in front of a machine than in front of a person, as the fear of negative judgment from others can lead to conformity.[32] While we often conform to human judgment to avoid scrutiny, we do not tend to be conformist or shy towards a machine, its impersonality offering a different kind of freedom and allowing us to push boundaries without the fear of social approval.

The issue of "producing content for the machine" touches on the broader cultural and social acceptance of our evolving, quasi-personal relationship with technology. As we have attempted to argue, this issue can be framed negatively - as fraught with economic and social dangers and potential degradation of our creativity - or more neutrally, as a stage in the ongoing technological and cultural evolution of humanity and the relationship between individuals and their devices. On the one hand, the technology we use becomes a part of us. By integrating with it, technology changes the way we think, act, and perceive the world. Conversely, this relationship is reciprocal: technology absorbs elements of our creativity and adapts in response. In the context of our discussion, the creative subject—whether a designer, writer, or artist—serves as an

intermediary and a catalyst for the machine, fostering an increasingly close dialogue where human and machine integrate, converge, and ultimately, fuse.

Notes

[1] Dennett, D. (1988). When philosophers encounter AI. *Daedalus: Proceedings of the American Academy of Arts and Sciences, 117*(1), 283–295. Reprinted in *Brainchildren: Essays on designing minds.* MIT Press, 1998.

[2] Whiddington, R. (2023). A Photographer Submitted an A.I.-Generated Image to a Prestigious Art Competition to Be 'Cheeky.' It Won a Top Prize Anyway. https://news.artnet.com/art-world/boris-eldagsen-photography-award-sony-ai-generated-images-dall-e-2286622.

[3] Schrader, A. (2023). A Photographer Wins a Top Prize in an A.I. Competition for His Non-A.I. Image. https://news.artnet.com/art-world/miles-astray-non-ai-photo-wins-ai-competition-2500809.

[4] Schrader, A. (2023), An Australian Photographer Was Disqualified From a Photo Contest After Her Submission Was Mistakenly Deemed A.I.-Generated. https://news.artnet.com/art-world/australian-photographer-disqualified-ai-generated-2337906: "Anderson, who regularly spots errors in photographs before printing them, said he did not see any imperfections in Dougherty's photo".

[5] Or, alternatively, to persuade an human to do something for the program, as in the 2023 case, when an AI chatbot tricked a TaskRabbit worker to solve a CAPTCHA by telling the worker it was not a robot and had impaired vision: Hurler, K. (2023). Chat-GPT Pretended to Be Blind and Tricked a Human Into Solving a CAPTCHA. https://gizmodo.com/gpt4-open-ai-chatbot-task-rabbit-chatgpt-1850227471.

[6] On the difficulty by humans to generate true random sequences, see https://calmcode.io/blog/inverse-turing-test

[7] Laarhoven, T., & Ponukumati, A. (2023). Towards transparent cheat detection in online chess: An application of human and computer decision-making preferences. In C. Browne, A. Kishimoto, & J. Schaeffer (Eds.), *Computers and Games. CG 2022. Lecture Notes in Computer Science, 13865*. Springer.

[8] Most "AI-originality" tests are actually "fake image" detectors, namely systems that are developed to test if an image has been artificially generated. See Ojha, U., Li, Y., & Lee, Y. J. (2023). Towards universal fake image detectors that generalize across generative models. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,

24480–24489; Sha, Z., Li, Z., Yu, N., & Zhang, Y. (2022). DE-FAKE: Detection and attribution of fake images generated by text-to-image generation models. https://arxiv.org/abs/2210.06998.

[9] Bell, P., & Offert, F. (2021). Reflections on connoisseurship and computer vision. *Journal of Art Historiography,* 24; Zanardelli, M., Guerrini, F., Leonardi, R., et al. (2023). Image forgery detection: A survey of recent deep-learning approaches. *Multimedia Tools and Applications, 82*, 17521–17566.

[10] Eastman, D. (2023). Data dignity: Developers must solve the AI attribution problem. https://thenewstack.io/data-dignity-developers-must-solve-the-ai-attribution-problem.

[11] Elran, S. R., & Zoran, A. R. (2024). Probabilistic craft: Materialization of generated images using digital and traditional craft. *Proceedings of the ACM on Computer Graphics and Interactive Techniques, 7*(4), Article 63. https://doi.org/10.1145/3664209.

[12] Dance, C. J., Ipser, A., & Simner, J. (2022). The prevalence of aphantasia (imagery weakness) in the general population. *Consciousness and Cognition, 97*, 103243.

[13] Zeman, A. (2024). Aphantasia and hyperphantasia: Exploring imagery vividness extremes. *Trends in Cognitive Sciences,* 28.

[14] Pykes, K. (2023). How to use Midjourney: A comprehensive guide to AI-generated artwork creation. https://www.datacamp.com/tutorial/how-to-use-midjourney-a-comprehensive-guide-to-ai-generated-artwork-creation.

[15] "Work of art are symbolic expressions, and […] they embody their meanings. The task of criticism is to identify the meanings and explain the mode of their embodiment". Danto, A. C. (1992). The art world revisited. In *Beyond the Brillo Box: The Visual Arts in Post-Historical Perspective.* University of California Press.

[16] Costello, D. (2007). Kant after Lewitt: Towards an aesthetics of conceptual art. In P. Goldie & E. Schellenks (Eds.), *Philosophy and Conceptual Art* (pp. 92-115). Oxford University Press.

[17] Adilova, A., & Shamoi, P. (2024). Aesthetic preference prediction in interior design: Fuzzy approach. *ArXiv*, abs/2401.17710.

[18] Murray, N., Marchesotti, L., & Perronnin, F. (2012). AVA: A large-scale database for aesthetic visual analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2408–2415.

[19] Abe, Y., Daikoku, T., & Kuniyoshi, Y. (2024). Assessing the aesthetic evaluation capabilities of GPT-4 with vision: Insights from group and individual assessments. *ArXiv*, abs/2403.03594.

[20] Frascaroli, J., Leder, H., Brattico, E., & Van de Cruys, S. (2024). Aesthetics and predictive processing: Grounds and prospects of a fruitful encounter. *Philosophical Transactions of the Royal Society B, 379*, 20220410. http://doi.org/10.1098/rstb.2022.0410.

[21] Muth, C., & Carbon, C. C. (2024). Predicting instabilities: An embodied perspective on unstable experiences with art and design. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 379*(1895), 20220416. https://doi.org/10.1098/rstb.2022.0416; Berlyne, D. E. (1960). *Conflict, arousal, and curiosity.* New York: McGraw-Hill.

[22] Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances, 10*(28). https://doi.org/10.1126/sciadv.adn5290.

[23] Loder, D. (2021). The aesthetics of digital intimacy: Resisting Airbnb's datafication of the interior. *Interiors, 11*(2–3), 282–308. https://doi.org/10.1080/20419112.2021.1945816; Halawa, M., & Parescoli, F. (2021). *Global Brooklyn.* Bloomsbury Publishing; Chayka, K. (2024). *Filterworld: How algorithms flattened culture.* Knopf Doubleday.

[24] Ren, J., Shen, X., Lin, Z., Mech, R., & Foran, D. J. (2017). Personalized image aesthetics. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 638-647).

[25] Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A., Babaei, H. R., LeJeune, D., Siahkoohi, A., & Baraniuk, R. (2023). Self-consuming generative models go MAD. *ArXiv*, abs/2307.01850.

[26] Fried, I., & Rosenberg, S. (2023). AI could choke on its own exhaust as it fills the web. https://www.axios.com/2023/08/28/ai-content-flood-model-collapse

[27] Valyaeva, A. (2023). People Are Creating an Average of 34 Million Images Per Day. https://journal.everypixel.com/ai-image-statistics.

[28] Kirschenbaum, M. (2023). Prepare for the Textpocalypse. *The Atlantic.* https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318.

[29] "To put that number in perspective, the total number of albums released per year in the 1980s, during the era of CDs and cassettes, was a few thousand": Rosenblatt, B. (2023). Spotify's New Royalty Model Confronts The Overabundance Of Music. https://www.forbes.com/sites/billrosenblatt/2023/10/26/spotifys-new-royalty-model-contends-with-the-age-of-musical-overabundance/.

[30] Helliwell, A. (2019). Can AI mind be extended? *Eventual Aesthetics, 8*, 93-120.

[31] Bendig, E., Erb, B., Schulze-Thüsing, L., & Baumeister, H. (2019). The next generation: Chatbots in clinical psychology and psychotherapy to foster mental health – A scoping review. *Verhaltenstherapie*. https://doi.org/10.1159/000501812.

[32] Steffen, L. (2019). This New App Lets Your Phone Judge Your Fashion Sense. https://www.intelligentliving.co/app-lets-phone-judge-your-fashion-sense/.