

# **AI Art and Artists: What They Are, What They Could Be, What They Should Be**

Dominic McIver Lopes

# AI Art and Artists: What They Are, What They Could Be, What They Should Be

Dominic McIver Lopes<sup>1,\*</sup> 

<sup>1</sup>University of British Columbia, Vancouver, Canada

\*Corresponding author: Department of Philosophy, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada.  
Email: [dom.lopes@icloud.com](mailto:dom.lopes@icloud.com)

## ABSTRACT

Generative AI systems provoke important practical questions, but they also provoke existential anxieties, which arise in response either to the reality or the possibility of AI art made by genuinely creative AI artists. This article first reviews recent arguments that existing AI systems do not make art and are not genuinely creative. It then argues that existential anxieties are not warranted if possible AI artists are just like human artists. Finally, it argues that, if they are not just like human artists, then we should design them to benefit us by satisfying an interest in aesthetic diversity. The upshot is that no existential anxieties are justified. In making this argument, this article draws lessons about AI art and artists and also about our values and interests.

ChatGPT, Midjourney, and Udio generate stories, images, and songs that amaze us. They can trouble us too, provoking existential anxieties. Existential anxieties are anxieties about affirmative answers to question such as: Are they art? Could there be AI artists? Would they be creative? Might they surpass us? Will they, as Lindsay Brainard asks, “challenge the special status we’ve taken ourselves to have by besting us in the domains we cherish most?” (2025, 1133). Anxieties that drive us to philosophy often persist despite any assurances that it might offer, and we will no doubt continue to worry as long as AI technologies remain unfamiliar and uncanny. Philosophy can only settle whether we have reason to worry. The first section below argues that the existential anxieties are not justified given the current state of the technology. The second section argues that we have no reason to worry about future possible AI art and artists, were they to turn out just like us. Maybe they will not be like us. In that case, the final section proposes that we should undertake a normative inquiry where we ask what they should be. Such an inquiry replaces existential worries with practical ones. Therefore, no existential anxiety is warranted. In the course of the argument, we can draw some lessons about AI art and artists and also about our values and interests.

## I. WHAT THEY ARE

Generative AI technology already raises thorny practical problems. It threatens to put some writers, image-makers, and composers out of business, it requires adjustments to copyright law, it accelerates an erosion of trust in media, and it amplifies social biases (Fallis 2021; Johnson 2021; Rini and Cohen 2022; Shoemaker 2024; Fazelpour and Magnani 2025; Hyska 2025). Sensitive to these practical problems,

Received: August 2, 2025. Revised: December 13, 2025. Accepted: December 29, 2025

© The Author(s) (2026). Published by Oxford University Press on behalf of the American Society for Aesthetics.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [reprints@oup.com](mailto:reprints@oup.com) for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

philosophers have recently sought to clarify what the technology can do. In clarifying its limitations, they propose necessary conditions on art and artistic agency, observe that existing technologies fail to meet the conditions, and conclude that there are no genuine AI artists or artworks. If this conclusion is correct, existential anxieties are not justified by the current state of the technology. A closer look at the reasoning for the conclusion will equip us to address anxieties about future or possible AI art and artists.

Philosophical and public attention focuses on generative AI systems with three notable features. First, their performance reflects a process of learning patterns in huge datasets. Having learned the patterns, they can iterate and combine them to generate novel outputs. Second, the novel outputs are no more predictable than those of human beings issued similar instructions. Illustrators commissioned to draw “Immanuel Kant eating a hot dog while standing at a lectern in front of a blackboard inscribed with the words ‘dogmatic slumber’” are unlikely to show a glass and concrete room with a panda in the audience. DALL-E meets similar expectations ([figure 1](#)). Moreover, the slightly perturbed facial expression, paneling details, open notebook, wire-rimmed specs, light fixtures, and espresso shot are more or less fitting, but they will not appear in most images made (by humans or computers) on the same prompt. Third, in order to generate novel outputs, generative AI systems such as DALL-E require prompts—user-generated descriptions for querying a network.

Do technologies with these three features meet necessary conditions on art and artistic agency? To begin with, are AI-generated stories, images, and music ever works of art? Philosophers have come nowhere near to consensus about what makes items works of art ([Lopes 2014](#)). An ambitious strategy is to consider a range of theories or families of theories of art ([Vlaad 2024](#)). A more modest strategy notes that, on all theories of art, nothing is a work of art unless it is an artifact, hence a product of an act, hence made intentionally. In the literature, it is thought obvious that generative AI systems are not agents because they do not have intentions ([Wojtkiewicz 2023](#); [Young and Terrone 2024](#); [Nannicelli 2025](#); [Anscomb forthcoming](#)). As long as generative AI systems require human prompts in order to generate output, it follows that the output is a work of art only in as much as it is partly a product of a human act of prompting.

The question about AI art raises a further question about the role that generative AI plays in making art. Since artists are agents who make art intentionally, generative AI systems are not artists if they have no intentions. Grant that they have no intentions (compare [Yee, unpublished manuscript](#)). It follows that they are not artists. Then are they tools in artists’ hands? Are they in the same league as paintbrushes or spell checkers? Following E. J. Lowe (2014), [Nick Young and Enrico Terrone \(2024\)](#) distinguish

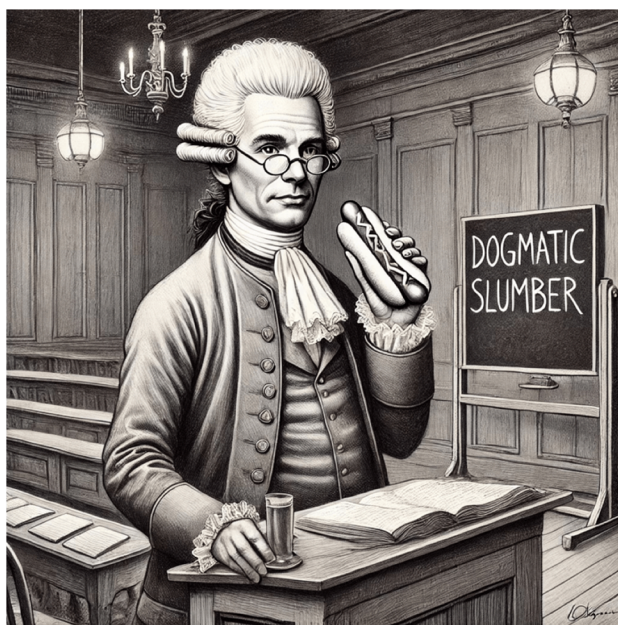


Figure 1. Dominic McIver Lopes, DALL-E-3 Image (2025) CC BY-SA.

two types of tools. Utensils, such as spoons and paintbrushes, require more or less continuous control on the part of their users in order to perform their functions. Machines, such as clocks and spell checkers, perform their functions without continuous user control, but they function well only when they are relatively predictable in their performance. DALL-E is obviously not an artistic utensil. Neither is it an art machine. After all, it is no more predictable than a human illustrator issued a similar prompt (and the human is not a machine). So, what is DALL-E, if neither an artist nor a tool in either of Lowe's two senses? Here answers diverge in productive and interesting ways.

One approach conceives the prompt-writing artist as producing an artwork in something like a collaboration with the generative AI system (for example, [Carroll 2025](#)). [Claire Anscomb \(2025\)](#) surveys several models of collaboration, which require equal collaborators to have, for example, shared intentions or mutual understanding. On these models, generative AI systems do not collaborate equally with artists, for the same reason that they do not make art. What is needed is a model of the AI system's contribution to joint activity that places it between being a tool and being an equal collaborator, where lack of predictability is a benefit. [René Mogensen \(2022\)](#), [Kathryn Wojtkiewicz \(2023\)](#), and [Anscomb \(2025\)](#) propose a division of labor between an artist, who provides executive agency, setting goals and monitoring results, and an assistive agency that does not set goals or monitor results but that uses intelligence to produce original and not entirely predictable results. The same division of labor is familiar in arrangements between people, including those in artists' studios.

Another approach is proposed by [Young and Terrone \(2024\)](#) and [Anthony Cross \(2025\)](#). Young and Terrone draw an analogy with gardening. A gardener designs a planting scheme and then tends the garden, coaxing and pruning in accordance with the scheme. Nonetheless, the garden is a piece of nature that operates autonomously and relatively unpredictably. Knowing this, good gardeners cede some control to nature, leaving room for the garden to grow in its own way within the parameters they set. In doing so, they allow for surprises. Likewise, an AI system can be used as an autonomous and unpredictable source of surprise. Making a similar point, Cross discusses artistic uses of AI systems as opportunities for artistic exploration. For example, in her *Critically Extant* series of 2022, artist Sofia Crespo prompts an AI image generator to make images of endangered or extinct species. Since the system's training exposed it to few images of these species, it "hallucinates" their appearance. The hallucination of an endangered fish poignantly exposes a blind spot in the AI's vision of the world, and therefore our vision too. An artist can make art by exploring the autonomous and unpredictable operation of a generative AI system (see also [Lopes 2009](#)).

To the questions about AI art and artists we can add a third. Are generative AI systems creative (see [Moruzzi \[2025\]](#) for an overview)? The answer is a foregone conclusion if creativity entails agency. On some theories, it does not. For example, [Margaret Boden \(2005\)](#) proposes that a system is creative just when its products are novel and valuable. Generative AI systems are creative, on this theory, because they generate novel and valuable stories, images, and music. However, Berys Gaut argues that Boden's theory overgeneralizes. It mistakenly counts as creative a "tree that produces an elegant and distinctive canopy of leaves" (2018, 129). Most theories of creativity add an agency condition to the novelty and value conditions. With an agency condition in place, we have the forgone conclusion. Generative AI systems are not creative as long as creativity implies agency in making: agency in making implies making intentionally, and the systems have no intentions. Apparently, we can learn no more by examining the possibility of creativity in generative AI than we already know from having examined its prospects for artistic agency.

This reasoning is a tad cheap. If creativity requires agency, then there is arguably more to the agency that it requires than making intentionally. I intend to bake a cake, the result is novel and tasty, but I exercise no creativity, because I am a hack, bereft of skill and experience. There is more to being creative than intentionally making something new and good. So, even if they never act intentionally, it would be significant if generative AI systems have what it takes to fill the gap.

Drawing upon [Gaut's \(2018\)](#) theory of creativity, [Anscomb \(2022\)](#) proposes that know how or skill fills the gap. To be creative, one must act intentionally but also out of skill. Therefore, she asks whether the novel and valuable outputs of generative AI ever result from the system's possessing know how or skill (see also [Brainard 2025](#)). Making through skill has some interesting features. When baked skillfully, a cake is not an accident, yet it is not a result of simply following a mechanical procedure. Creativity is never formulaic and never accidental. [Maria Kronfeldner \(2018\)](#) proposes that the creative use of skill

is spontaneous in the sense that it does not involve fixing on a precise end and a precise means to the end. A creative baker substitutes reliance on skill for reliance on a determinate end or a set of predetermined means. Finally, skill must be guided by evaluations: a skilled baker monitors how well things are going.

How does generative AI stack up against the skilled baker? On the plus side, its outputs are neither accidental nor formulaic. Relatedly, it has neither a precise end in mind nor a precise plan for achieving that end. On the minus side, its process is not guided by evaluations. In effect, it throws everything against the wall to see what sticks with users.

Brainard (2025) adds another condition on creative agency. An agent's creativity must be to their credit as an agent. To begin with, it must be a merit in an agent. According to Brainard, creativity is a merit in an agent because curiosity is a merit in an agent and creativity implicates curiosity. Since it relies on skill in order to get novel and valuable results, "creativity involves a sort of exploration—a pursuit of something that is, in some sense, unknown" (Brainard 2025, 1139). The appeal to the value of curiosity to explain why creativity is a merit in an agent can also explain why their creativity is to their credit. Brainard argues that, at the very least, to take an agent's merits as being to their credit is to appraise their merits as an adopter of ends (2025, 1145–46 after Watson 1996). An adopter of ends must be curious about ends. For example, to take Cindy Sherman's creativity in melding performance art with photography as being to her credit is to appraise her as an adopter of artistic ends. In part, it is to see her as exploring, through art-making, what her ends might be.

Generative AI systems such as DALL-E are not curious in this way (even if Young and Terrone and Cross are right about their affording acts exploration by artists). After all, their ends are provided externally, through prompts. They are not creative in the sense that being creative is to the credit of a curious, end-adopting agent.

In short, existing AI art is made under human oversight, and no existing generative AI technologies meet all necessary conditions on being an artist or being creative. Hence, existential anxieties are not justified by the current state of AI technology. Obviously, though, the technology might evolve to meet the necessary conditions on creative artistic agency. Building upon this discussion of intentionality, executive agency and creativity, the next two sections consider possible AI technologies of the future.

## II. WHAT THEY COULD BE

The space of possibility, even practical possibility, is huge. Here is a handy partition. Either possible AI artists are very much like us, or they are not. This section considers the first disjunct, where possible AI artists are very much like us and their art is therefore very much like ours. Suppose that they could exhibit intentionality and executive agency. Suppose that they could explore with curiosity. Suppose they could be just like us in all the ways that matter. Does the possibility warrant existential anxiety? In order to answer this question, we need a procedure to explore the possibility, and we need that procedure to clarify "the ways that matter." The following procedure uses David Chalmers's simulation hypothesis to identify what matters and to argue that the possibility warrants no existential anxiety.

According to the simulation hypothesis: "we are and always have been in an artificially designed computer simulation" (Chalmers 2022, 29; see also Bostrom 2003). Some simulations are local; they simulate a proper part of a whole, such as a tropical cyclone, which is part of an oceanic weather system. Chalmers's simulation is global, simulating a total world. Some simulations are temporary; they simulate what happens over a limited spell of time. Chalmers's simulation is lifelong. Let a sim be any item in a simulation. Biosims are biological creatures connected to a computer one world up, undergoing experiences of being in our world (think Neo in *The Matrix*). Unlike a biosim, a pure sim is not a state of a biological creature one world up; it is a subroutine run on the computer one world up (think NPCs). In Chalmers's simulation, we are all pure sims rather than biosims. In brief, the hypothesis is that everything in our world is and always has been nothing more than part of a state of an artificially designed computer.

Generative AI suggests a relatively direct way to implement a permanent, global simulation comprising all pure sims. A system's neural networks are trained to detect physical, biological, psychological, and social patterns one world up. Once trained, the system can iterate and combine to generate new patterns, which are the goings on in our world. A relatively indirect implementation sets the system up

to perfectly simulate the laws of physics. Some of the simulations run in accordance with those laws can also simulate biological, psychological, and social phenomena. Two conditions must be satisfied. The simulation of the laws must be sufficiently complete—the rather partial simulation of the laws of motion in the 1972 video game *Pong* fails to simulate much else (it does not even simulate tennis). Moreover, biological, psychological, and social phenomena must supervene on physical processes. So long as everything supervenes on the physical, simulating the physical thereby simulates everything else.

Perhaps the simulation hypothesis strikes you as fanciful. No matter, for all we need is that the hypothesis is possible. Chalmers recruits its possibility to shed light on several issues in philosophy, and perhaps its possibility can also help us to assess existential anxiety about possible AI artists who are very much like us and make art very much like ours. To assess this anxiety, we need to suppose that they are like us in all the ways that matter. On the simulation hypothesis, we are them. If we are them, then they are like us in all the ways that matter. Working with the simulation hypothesis might bring out what matters and thereby enable us to assess existential anxiety about possible AI artists who are very much like us—that is, equivalently, about how to think about art and artists if the simulation hypothesis is true.

Here is a little argument:

1. pure sims are real,
2. so, if the simulation hypothesis is true, then the entities around us are real;
3. the entities around us include artworks and artists,
4. so, if the simulation hypothesis is true, then the artworks and artists around us are real.

Suppose that the simulation hypothesis is true. In that case, everything in our world is a pure sim. Everything in our world is a digital entity, an informational structure. The constituents of our world include Ian Rankin and his novel, *Fleshmarket Close*, Cindy Sherman and her photograph, *Untitled Film Still #3*, and Jon Batiste and his song, “Freedom.” So, all six are pure sims. Among the pure sims are sim art and sim artists—call them “art+” and “artists+.” The little argument concludes that art+ and artists+ are real.

Is the little argument sound? What does it tell us about what matters? What does it imply for existential anxiety about art+ and artists+?

The way into all three questions is to begin with the first one. Premise (2) follows from (1) plus the definition of a pure sim. Premise (3) is a true empirical observation; indeed, it is one upon which the existential anxieties depend. The conclusion follows directly from (2) and (3). Therefore, the crux is (1), the commitment to digital realism. Is it true? Why think that pure sims are real?

Chalmers floats five conditions, each of which suffices for an entity to be real (2022, chap. 6). First, it is real if it exists, where its being perceivable and measurable is a good indication that it exists. Second, it is real if it is causally impacted or impactful. Third, it is real if it does not depend on anyone’s mental states. Fourth, it is real if it is not illusory, if it is roughly as we believe it to be. Finally, it is real if it is genuine, not ersatz or fake.

Art+ and artists+ arguably meet all five conditions. Suppose that the simulation hypothesis is true. All the same, “Freedom” exists: I have heard it. It has causal impact: listening to it consoled me as much as anything could on the morning of Wednesday, November 6, 2024. That Batiste surprised me with his cover of “Für Elise” shows that his doings do not depend on my expectations. Most of our beliefs about Batiste and “Freedom” are accurate. Finally, some artists are phony; some art works are not the real deal. The distinction between the phony and the fake, on the one hand, and the genuine, on the other hand, obtains within the simulation that is our world. Batiste and “Freedom” are on the kosher side of the line.

This reasoning is admittedly too quick to seriously engage the existential worries. Our thinking about what is real had better engage the worries as ones that arise in specific relation to art and artists. Whatever it is to be real, that conception of reality must vindicate the reality, in our appreciation of them as artists and artworks, of Rankin, Sherman, Batiste, *Fleshmarket Close*, *Untitled Film Still #3*, and “Freedom.” In addition, it had better vindicate the reality of these artists’ images of themselves as creative. It will not do for them to be real in a way that matters not at all to their self-images or to our engagement with them as artists and works of art.

As a way to work toward the relevant conception of being real, here are three objections to the too quick reasoning to the claim that art+ and artists+ are real. First, if the simulation hypothesis is true, then it turns out that artists+ and art+ are not as we believe them to be (or as they believe themselves to be). We (and they) believe them to be biological creatures, and we believe that we engage with artworks by interacting with something material. Second, in simulations, causal interactions turn out to be information processing, but the causal interactions in which artists and art are involved are not transitions of informational states in accordance with algorithms. Finally, art+ and artists+ are not the real McCoy because origins matter, and they do not have the right origins. In short, art+ and artists+ are not real in three respects that matter to our engagement with art and artists.

Start with the second objection, targeting causality, because the reply gets to the heart of things. When it comes to art, what matters are aesthetic responses. [Robbie Kubala \(2024\)](#) helpfully disentangles three modes of response. In aesthetic judgment, we make a determination as to a work's aesthetic value. In aesthetic appreciation, we see the work as aesthetically good, partly by seeing the features that make it good. In aesthetic understanding, we grasp how those features make it good; we grasp how they ground its goodness. Now, aesthetic understanding involves interpretation. Part of what goes into aesthetic understanding is an interpretation of a work as made by an artist in a context of production for a context of reception. For example, an aesthetic understanding of *Untitled Film Still #3* must take on board its having been made by Sherman in 1977 as a feminist critique of the depiction of women in Hollywood movies, as well as instigating the development of a new kind of photography where scenes are carefully staged before being shot ([Lopes 2018](#), 48–65). Likewise, an aesthetic understanding of Sherman as an artist must take on board how her artistic project was to meld performance art with photography, infusing images with, as Arthur Danto puts it, “the promise and threats of the performance artist's real presence” (1987, 136). Sherman will express her image of herself as creative by telling a similar story of reaction, influence, and impact. Notice that interpretive aesthetic understanding attributes causes and effects. We (and she) make sense of *Untitled Film Still #3* as a product of her project in a social context, impacting others in the context. We (and she) make sense of her doing something that shaped her subsequent work and the work of others. Here is the takeaway: all the causal power we need to attribute to art+ and artists+ is what we attribute in the project of interpretive aesthetic understanding. Art+ and artists+ are real in the sense that we interpret them as impacted by and having an impact on other art+ and other artists+.

Granted, Sherman is an artist+ and her photograph is art+ only in so far as each occupies a place in a flow of information. Neither is causal because material; both are causal because they are involved in transitions of informational states in accordance with algorithms. By the same token, however, Sherman is an artist and her photograph is art only in so far as neurons are firing and photons being absorbed. Just as nothing happens in reality except through microphysical processes, nothing happens in reality+ except through informational processes. The objection needs to be that what is happening at the more basic level (of bits or its) makes a difference at the level of interpretive aesthetic understanding. What difference is that? Sherman's influence on, say, Jeff Wall, is the same whether she is composed of atomic or informational structures. The heart of our engagement with art and artists (and the heart of artists' creative self-image) is an interpretive activity that applies equally to art and art+ and to artists and artists+. All are real in the sense that they are subject to (and subject themselves to) causal interpretation.

With this result in hand, we can make quick work of the objection from illusion. That objection was that art+ and artists+ are not what we believe them to be, for we believe them to be fundamentally material, whereas they are fundamentally informational. Likewise, artists+ are not what they believe themselves to be, fundamentally material rather than informational beings. These are indeed mistakes if the simulation hypothesis is true, but the question is whether they are mistakes that matter. What matters in our engagement with art and art+, in our stance toward artists and artists+, and in the self-image of artists and artists+ is the same: the activity of interpretive aesthetic understanding. Artists+ and art+ afford interpretive aesthetic understanding even if our (and their) beliefs about their fundamental make up are false.

That leaves the third objection, that art+ and artists+ are not real because they are not genuine art or artists. The objection can be amplified by drawing upon the replies to the first two objections. Interpretive aesthetic understanding extends to the deepest origins of its objects. A “Rolex” watch offered for

sale on eBay might not differ at all from a genuine Rolex with respect to materials or quality of workmanship. The difference between them is a difference in origins or provenance, and that difference matters for some purposes. It matters for art, in particular. A perfect copy of *Untitled Film Still #3* affords interpretive aesthetic understanding whose content, when correct, is not the same as the content of the correct interpretive aesthetic understanding of Sherman's original (Dutton 1985). The copy does not look as it does because it was made by Sherman in 1977 in sympathy with feminist critiques of images of women in movies and with trends in the avant-garde. Instead, it looks as it does because, say, its maker wanted to exploit the market value of the Sherman *oeuvre*. Origin matters when it makes a difference to interpretive understanding.

Here are two origin stories about Sherman and her photograph. Origin Story 1 tells of the 1970s art scene and how it was affected by her work. Origin Story 2 tells of the fundamental constituents of the world: it tells of its, not bits. Origin Story 1 is correct whether Sherman is an artist or an artist+ and whether *Untitled Film Still #3* is art or art+. The truth of Origin Story 2 entails the falsity of the simulation hypothesis. However, it is only Origin Story 1 that pertains to interpretive aesthetic understanding: it is not relevant whether Origin Story 2 is correct. Origin matters only when it makes a difference to interpretive aesthetic understanding, and Origin Story 2 is not relevant to interpretive aesthetic understanding. In short, an item's basic make up (its or bits) does not make a difference to the reality of art+ and artists+.

Here is the argument for the conclusion that we have no reason to be anxious about the possibility that art and artist will turn out to be just like us. First, we have reason to be anxious about this possibility only if we have reason to be anxious about the possibility that we just are them. This premise follows from the symmetry of similarity and identity. Second, we have reason to be anxious about the possibility that we just are them only if the discovery that we are them would make a difference to our appreciations or our creative self-images. In other words, we only have reason to be anxious about what matters. Third, the moral of the simulation hypothesis is that the discovery that we are them would make no difference to our appreciations or our images of ourselves as creative. Rankin, Sherman, and Batiste are artists—or maybe they are artists+. *Fleshmarket Close*, *Untitled Film Still #3*, and “Freedom” are art—or maybe they are art+. Were we to discover that the simulation hypothesis is true, our practices give us no reason to take back what we have thought of their talents or the beauty and significance of their work. They give Sherman no reason to question whether she was trying to convey a feminist message. What we do with art+ in light of its being made by artists+ is just what we do with art in light of its being made by artists. From these premises, it follows that we have no reason to be anxious about the possibility that we just are them and hence that we have no reason to be anxious about the possibility that art+ and artists+ will turn out to be just like us.

### III. WHAT THEY SHOULD BE

Either possible AI artists are very much like us, or they are not. The second disjunct remains. Let an artist± be an AI system that satisfies the necessary conditions on artistic agency and creativity, as set out in the first section, and let the products of their creative agency be works of art±. In addition, let artists± and art± differ from artists and art in ways that matter—that is, in ways that impact interpretive aesthetic understanding, as characterized in the previous section. Intuitively, we have reason to worry that we might share a world with beings who have the kind of agency and creativity that artists have but who are not just like us. They might harm us aesthetically or they might benefit us. Either way, they will have been products of our design decisions. Therefore, the question to ask is, how should we build them? What should they be?

The literature on what has come to be called “value alignment” assumes that we should build AI systems that benefit us by designing them so that their interests align with our interests (for example, Bostrom 2014). An AI system's interests are what are represented by a function by means of which it evaluates options for what to do (for example, Peterson 2019, 20). The idea is that we should design AI systems to evaluate options much as we would do. In particular, Martin Peterson (2019) proposes the principle that, at each point in time, we should design AI systems whose interests align with some of our interests at that point in time. Alice C. Helliwell (2024) adapts Peterson's principle to the design of artists±. According to the,

aesthetic alignment principle: we should design artists $\pm$  to benefit humans aesthetically, and at each point in time the best way to do this is to align the aesthetic interests of artists $\pm$  with some of the aesthetic interests that we embrace at that time.

The quantifier in the principle's final clause is license to build artists $\pm$  who do not "take all [their] cues from current human values" (Helliwell 2024, 7).

The principle is incomplete. It mandates an alignment of aesthetic interests while saying no more about them than that they are aesthetic. In order to build AI systems whose aesthetic interests align with ours, we need to supplement the principle with a conception of our aesthetic interests.

Implicit in Helliwell's case for the principle is a conception of our aesthetic interests: they are interests in what we do or will appreciate (2024, 15–16). Thus, she writes that "AI might surprise us with something we did not expect to appreciate" (2024, 15). For example, generative AI image systems make weird images that fascinate us (for example, some show hands with too many fingers). Who would have guessed? Moreover, what we appreciate changes over time, and the initial reception of painters such as Van Gogh and Picasso shows how poorly past appreciative preferences predict future ones. We now get the benefit of appreciating Van Gogh and Picasso only because they discounted their contemporaries' appreciative preferences. Therefore, in order to get the future benefit of appreciating something that we do not appreciate at the current time, we ensure that young artists do not defer too much to our current appreciative preferences. By the same reasoning, we should build artists $\pm$  whose aesthetic interests align with some of our appreciative preferences at a time, just as the aesthetic alignment principle suggests. So designed, they are free to produce art $\pm$  that we do not appreciate but will one day come to appreciate; they will produce art $\pm$  that is "ahead of its time," as Helliwell puts it (2024, 16).

Consider some claims about interests in appreciation. Helliwell's argument obviously implies that,

(c) our aesthetic interests include satisfying our current appreciative preferences.

Arguably, however, it requires more:

(c + F) our aesthetic interests include satisfying our current and future appreciative preferences.

That we will one day appreciate some art $\pm$  that we do not now appreciate is no reason to design artists $\pm$  in line with the aesthetic alignment principle unless we have an interest in our future appreciative preferences. Surely such an interest is an aesthetic interest. Interpreted through (c + F), the aesthetic alignment principle says that we should build artists $\pm$  in tune with some of our current appreciative preferences: doing so will serve some of our future appreciative preferences, though we do not know yet what those preferences will be.

A claim stronger than (c + F) confines our aesthetic interests to current and future appreciative preferences. That is,

(s) our aesthetic interests consist in satisfying our current and future appreciative preferences.

However, the argument for the aesthetic alignment principle does not imply (s), and (s) is open to challenge. We might have aesthetic interests in items that we do not and will not appreciate. Others appreciate them, but not us, and we might have an interest in their appreciations. Interests not confined to what we do or will appreciate are, loosely speaking, non-partial:

(NP) our aesthetic interests include satisfying our current and future appreciative preferences and respecting others' current and future appreciative preferences.

Recent work on aesthetic value supports (NP), providing an account of respect. Three ideas are pertinent. Frank Sibley (1959, 1965) argued that aesthetic value is a determinable whose determinates are values such as being graceful, dynamic, or balanced. In other words, nothing is beautiful except in some determinate way, much as nothing is red unless it has some shade of red—say, scarlet or crimson. Kendall Walton (1970) added that the determinates of beauty are relative to kinds of art. A sequence of moves

that is dynamic in ballet is insipid in boogaloo and *kathakali*. Other sequences of moves make for dynamism in boogaloo and *kathakali*. Building on Sibley and Walton, [Dominic McIver Lopes \(2014, 2018\)](#) represents each kind of art as anchored in a social practice or culture. Ballet, boogaloo, and *kathakali* are cultures whose members coordinate with one another because they share a conception of the determinates of beauty. The world of ballet is one where choreographers, dancers, set designers, publicists, critics, and audiences share a conception of what sequences of moves are dynamic or insipid, frenzied or measured.

On this picture, first, a capacity to appreciate items in one aesthetic culture does not imply a capacity to appreciate them in arbitrarily any other aesthetic culture. Someone schooled in ballet is unlikely to get a great deal from a *kathakali* performance. Second, culture-constituting conceptions of beauty do not speak against each other. In some domains, such as ethics, value schemes do speak against each other. Those loyal to an ethics of non-violence cannot abide the honor code of the warrior, and vice versa. Conceptions of beauty are not like this. Ballet aficionados need not scorn the conceptions of beauty in boogaloo or *kathakali*; they are in a position to respect them, where “respect” names a pro-attitude to values that we do not and will not appreciate ([Lopes 2024](#), chap. 4 after [Raz 2004](#), 160–4).

Assume that when we are in a position to respect an aesthetic value, then we have aesthetic reason to respect it, and assume that our aesthetic reasons determine our aesthetic interests. It follows that we have an aesthetic interest in respecting other’s values, when they are ones that we do not and will not appreciate. To accommodate this broader conception of aesthetic interests, replace the “our” in (NP) with a reference to cultures:

(NP\*) for any two aesthetic cultures,  $K_1$  and  $K_2$ , the aesthetic interests of members of  $K_1$  include satisfying the current and future appreciative preferences of members of  $K_1$  and respecting the current and future appreciative preferences of members of  $K_2$ .

Our aesthetic interests according to (NP\*) are broader than those according to (C + F). The former includes an interest in respecting aesthetic values that exceed the limits of what we do or will appreciate.

Why accept (NP\*)? [Lopes \(2018, 2024\)](#) gives a raft of arguments that imply (NP\*), so a general argument for it might rely on his network theory of aesthetic value and his cosmopolitan theory of aesthetic injustice. However, for present purposes, it would be better to give an argument for (NP\*) that specifically addresses the question of what art± and artists± should be. To do this, we can argue for interpreting the aesthetic alignment principle, which specifically concerns artists±, through (NP\*). Interpreted through (NP\*), the aesthetic alignment principle states that the best way to build artists± who benefit us is to align their aesthetic interests with some of our current appreciative preferences: we can then expect their art± either to satisfy some of our present or future appreciative preferences or to realize values that we respect. What follows is a tenability argument: the objections to interpreting the value alignment principle through (NP\*) fail.

Granting that (NP\*) is true and that the aesthetic alignment principle is apt for designing artists±, one might still object to interpreting the principle through (NP\*). One might restrict (NP\*) to human beings or biological creatures. Ballet fans have an aesthetic interest in respecting *kathakali* values, but they have no interest in respecting the aesthetic values of artists±, non-human machines.

Such a restriction would be ad hoc, though. (NP\*) identifies an interest in respecting aesthetic values that we are in a position to respect and yet do not and will not appreciate. Those who have the values need resemble us no more than is necessary for us to be in a position to respect their values.

This reply suggests a better objection to interpreting the aesthetic alignment principle through (NP\*). We have no interest in respecting aesthetic values in art± unless we have reason to respect the values, and we have no reason to respect the values unless we are in a position to respect them. We are in no such position. Artists± do not resemble us as much as is necessary for us to be in a position to respect their aesthetic values. Hence, we have no interest in respecting the aesthetic values of artists±.

The crux is what it takes for us to be in a position to respect the aesthetic values of artists±. How must artists± be like artists? Gathering all the lessons of this article, the proposal is this. First, artists± must meet the necessary conditions laid down in the opening section. Second, they must submit to interpretive aesthetic understanding, as characterized in the previous section. Third, as the aesthetic alignment principle provides, their aesthetic preferences must align with some of ours. Finally, as (NP\*) requires,

they must belong to aesthetic cultures. We can recognize creatures who satisfy these conditions as responding to aesthetic values: no further conditions need be met by artists± whose values we are in a position to respect. Accordingly, plans for building artists± should aim at creating social individuals who submit to interpretive aesthetic understanding. We vindicate interpreting the aesthetic alignment principle through (NP\*) if these plans are viable.

Since nothing indicates viability better than what we can already build, here is a case study of Ai-Da, an autonomous, robotic AI system (see <[www.ai-darobot.com](http://www.ai-darobot.com)>). The claim is not that she is an early artist±. Her designers do not say that she meets the necessary conditions laid down in the first section. She is certainly not a social individual. Nevertheless, she illustrates progress toward individuality and sociality, where each figures in interpretive aesthetic understanding.

Ai-Da is equipped with a suite of capacities, many quite different from those of generative AI systems such as Midjourney (figure 2). She is fitted with eyes that give her visual information about her surroundings, and that information is processed for her use. In particular, she wields a robotic arm to draw or paint what she sees. In doing so, she can tap memory of images she's made in the past, and she can access collective memory through databases and learning networks. Then she can talk about art, the images she makes, and herself—a topic on which she testified to the UK House of Lords. As artist-in-residence at the Ben Nicholson Studio in 2022, she responded to his work by making work of her own.

Start with individuality. With attention nowadays focused on generative AI, Ai-Da's embodiment is immediately striking, and embodiment is a good means to individuality. An embodied creature has a unique viewpoint on the world, a spatiotemporally constrained proper subset of worldly information. In addition, they act upon the world in ways that are constrained both by their spatiotemporal trajectory and by the kind of body that they have. All this feeds into egocentric memory, a building block of episodic memory.

Just as striking are Ai-Da's communicative capacities. At first glance, her communicative performance resembles that of systems like ChatGPT. Unlike ChatGPT's, though, her first-person utterances originate from and refer to an embodied and remembered perceptual and practical viewpoint. For this reason, her statements about herself ring less hollow than those of ChatGPT.

Ai-Da's embodiment, memory, and knack at communication invite interpretive aesthetic understanding. Audiences are most fascinated by her self-portraits (for example, figure 2), which are hard not to see as images she made of her body by looking at her body. Moreover, their characteristic jagged lines reveal the workings of her visuomotor system and the mechanics of her arm. In general, what she draws is a function of her journey through the world: making sense of her work in Nicholson's studio means knowing precisely that she worked in that place. In speaking about herself, she provides an insider's perspective that is to be reconciled with our perspective as outsiders. The reconciliation is typical of the project of interpretive aesthetic understanding.

Sociality requires memory and communication plus a capacity to recognize and act in accordance with shared interests. Recognizing shared appreciative preferences requires the capacity to classify in accordance with a conception of aesthetic value. AI systems handily classify works as “impressionist” or “pop,” for example.



Figure 2. Leemurtz, *Ai-Da with Self-Portrait* (2021) CC BY-SA.

Ai-Da recognizes art in her environment, responds to it by making images of her own, comments on what she has produced, and processes feedback about it from others. None of this occurs at a high level of sophistication. Her interaction with the social environment of art is not as rich as Sherman's. Ai-Da's rudimentary sociality invites only a rudimentary interpretive understanding of her activity as a product of interactions with others. All the same, she provides proof of concept (compare [Cascales 2023](#)).

We have a tenability argument for interpreting the aesthetic alignment principle through (NP\*). The objection was that we may so interpret the principle only if we have an interest in respecting aesthetic values in art $\pm$ , but we have such an interest only if we have reason to respect the values, and we have reason to respect the values only if we are in a position to respect them, yet artists $\pm$  do not resemble us as much as is necessary for us to be in a position to respect their aesthetic values. The reply is that we are in a position to respect artists $\pm$ ' aesthetic values when four conditions are met: artists $\pm$  act intentionally and creatively, submit to interpretive aesthetic understanding, align their aesthetic preferences with some of ours, and participate in aesthetic cultures as social individuals. Ai-Da does not meet all four conditions, but her existing capacities indicate potential for creatures like her to meet them. So, we have no reason to doubt that the aesthetic alignment principle can be interpreted through (NP\*).

The aesthetic alignment principle recommends building for difference, but it encourages building for greater difference when interpreted through (NP\*). Ai-Da's body makes an aesthetically significant difference to her work. Algorithms exist to nudge generative AI systems to make works that deviate from and thereby modify conceptions of aesthetic value on which they have been trained ([Elgammal et al. 2017](#), [Mazzone and Elgammal 2019](#)). According to (NP\*), we have reason not to minimize the differences between human and AI performance: we may explore and emphasize them. Sometimes, as Helliwell argues, we will benefit by coming to appreciate art $\pm$  in the future. However, our interest in respecting aesthetic difference means that we have reason to build artists $\pm$  whose art $\pm$  we do not and will not appreciate. Is that not an intriguing design proposition? If it is, then we have no reason for existential worries about possible AI artists who are not very much like us.

Perhaps existential anxiety is a symptom of enlightenment. Copernicus discovered that we are not the center of the universe, Charles Darwin that we are not enthroned atop the tree of life, and Sigmund Freud that we are not masters of our own selves. Their discoveries disorient some, even today. Geoffrey Hinton might deserve credit for breaking up the human monopoly on art. We are not yet there, but we should welcome the event, if AI technology presses us to identify our aesthetic interests and to advance them in ways we never expected.<sup>1</sup>

## REFERENCES

- Anscomb, Claire. 2022. "Creating Art with AI." *Odradek: Studies in Philosophy of Literature, Aesthetics, and New Media Theories* 8: 13–51.
- Anscomb, Claire. 2025. "AI: Artistic Collaborator?" *AI and Society* 40: 3419–29.
- Anscomb, Claire. Forthcoming. "Who Authors AI Art? (And Why Does It Matter?)." *AI and Society* (Forthcoming).
- Boden, Margaret. 2005. "Creativity." In *Routledge Companion to Aesthetics*, edited by Berys Gaut and Dominic McIver Lopes, 477–88, 2nd ed. London: Routledge.
- Bostrom, Nick. 2003. "Are You Living in a Computer Simulation?" *Philosophical Quarterly* 53: 243–55.
- Bostrom, Nick. 2014. *Superintelligence*. Oxford University Press.
- Brainard, Lindsay. 2025. "The Curious Case of Uncurious Creation." *Inquiry* 68: 1133–63.
- Carroll, Noël. 2025. "Are We Inventing Ourselves out of Our Own Usefulness? Striking a Balance between Creativity and AI." *AI and Society* 40: 249–51.
- Cascales, Raquel. 2023. "Interpreting AI-Generated Art: Arthur Danto's Perspective on Intention, Authorship, and Creative Traditions in the Age of Artificial Intelligence." *Polish Journal of Aesthetics* 71: 17–29.
- Chalmers, David. 2022. *Reality+: Virtual Worlds and the Problems of Philosophy*. New York: Norton.
- Cross, Anthony. 2025. "Tool, Collaborator, or Participant: AI and Artistic Agency." *British Journal of Aesthetics* 65: 535–51.
- Danto, Arthur C. 1987. "Cindy Sherman." *The Nation* 245: 134–7.
- Denis, Dutton, ed. 1985. *The Forger's Art: Forgery and the Philosophy of Art*. University of California Press.
- Elgammal, Ahmed Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. 2017. "CAN: Creative Adversarial Networks Generating 'Art' by Learning about Styles and Deviating from Style Norms." *ArXiv* 1706.07068v1.
- Fallis, Don. 2021. "The Epistemic Threat of Deepfakes." *Philosophy and Technology* 34: 623–43.
- Fazelpour, Sina, and Meica Magnani. 2025. "Aspirational Affordances of AI." *arXiv* 2504.15469.

- Gaut, Berys. 2018. "The Value of Creativity." *Creativity and Philosophy*, edited by Berys Gaut and Matthew Kieran. London: Routledge, 124–39.
- Helliwell, Alice C. 2024. "Aesthetic Value and the AI Alignment Problem." *Philosophy and Technology* 37: 1–21.
- Hyska, Megan. 2025. "The Politics of Past and Future: Synthetic Media, Showing, and Telling." *Philosophical Studies* 182: 137–58.
- Johnson, Gabbrielle M. 2021. "Algorithmic Bias: On the Implicit Biases of Social Technology." *Synthese* 198: 9941–61.
- Kronfeldner, Maria. 2018. "Explaining Creativity." *Creativity and Philosophy*, edited by Berys Gaut and Matthew Kieran, 213–29. New York: Routledge.
- Kubala, Robbie. 2024. "Varieties of Aesthetic Response." *Philosophical Topics* 5: 43–59.
- Lopes, Dominic McIver. 2009. *A Philosophy of Computer Art*. London: Routledge.
- Lopes, Dominic McIver. 2014. *Beyond Art*. Oxford University Press.
- Lopes, Dominic McIver. 2018. *Four Arts of Photography: An Essay in Philosophy*. Oxford: Wiley-Blackwell.
- Lopes, Dominic McIver. 2024. *Aesthetic Injustice*. Oxford University Press.
- Mazzone, Marian, and Ahmed Elgammal. 2019. "Art, Creativity, and the Potential of Artificial Intelligence." *Arts* 8: 26–9.
- Mogensen, René. 2022. "Computation, Creativity, and Improvised Music." *Odradek: Studies in Philosophy of Literature, Aesthetics, and New Media Theories* 8: 53–76.
- Moruzzi, Caterina. 2025. "Artificial Intelligence and Creativity." *Philosophy Compass* 20: e70030: 1–8.
- Nannicelli, Ted. 2025. "Mass AI-Art: A Moderately Skeptical Perspective." *Journal of Aesthetics and Art Criticism* 83: 343–57.
- Peterson, Martin. 2019. "The Value Alignment Problem: A Geometric Approach." *Ethics and Information Technology* 21: 19–28.
- Raz, Joseph. 2004. *Value, Respect, and Attachment*. Cambridge University Press.
- Rini, Regina, and Leah Cohen. 2022. "Deepfakes, Deep Harms." *Journal of Ethics and Social Philosophy* 22: 143–61.
- Shoemaker, Eric. 2024. "Is AI Art Theft? The Moral Foundations of Copyright Law in the Context of AI Image Generation." *Philosophy and Technology* 37: 1–21.
- Sibley, Frank. 1959. "Aesthetic Concepts." *Philosophical Review* 68: 421–50.
- Sibley, Frank. 1965. "Aesthetic and Nonaesthetic." *Philosophical Review* 74: 135–59.
- Vlaad, Sofie. 2024. "A Portrait of the Artist as a Young Algorithm." *Ethics and Information Technology* 26: 1–11.
- Walton, Kendall. 1970. "Categories of Art." *Philosophical Review* 79: 334–67.
- Wojtkiewicz, Kathryn. 2023. "How Do You Solve a Problem Like DALL-E 2?" *Journal of Aesthetics and Art Criticism* 81: 454–67.
- Yee, Adrian K. "Computers Can Author Art," unpublished manuscript.
- Young, Nick, and Enrico Terrone. 2024. "Growing the Image: Generative AI and the Medium of Gardening." *Philosophical Quarterly* 75: 310–19.

## ENDNOTES

- 1 My thanks to two anonymous referees, Adrian Yee, and audiences at the Jeffrey Rubinoff Company of Ideas Forum, the Mississippi Philosophy Association, the University of Campinas, and the Colégio Brasileiro de Estudos Avançados. A special thanks to Aidan Meller and Lucy Seal for discussions of their work on Ai-Da.

Funding: none declared.

Conflicts of interest: None declared.