

Art, Authorship, and Artificial Intelligence

By

Sofie Shannon Vlaad

A thesis submitted to the Department of Philosophy

in conformity with the requirements for the

Degree of Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

April, 2026

Copyright © Sofie Vlaad, 2026

Abstract

Scholars have recently taken interest in the phenomenon of AI art. I engage a variety of questions surrounding this topic, especially those pertaining to the art status of AI-generated work and how we might adjudicate authorship in such cases. I begin by arguing that according to contemporary accounts of art, some AI-generated images can count as artworks. I contend that it makes more sense to conceive of such popular works of AI art as bad art rather than non-art. Next, I argue that some AI-generated artworks do possess unique aesthetic properties. I show that by deploying a technique I call bias-prompting, artists are able to exploit the algorithmic bias of AI systems to generate artworks that possess an aesthetic feature I call misalignment. I then consider whether AI is an artistic medium or artform and argue that AI is the vehicle through which the artistic medium and artform of AI art is expressed. I conclude by suggesting that in the case of AI-generated texts there is a curator in place of an author. Whereas an author creates what I call Traditional Texts, a curator discovers what I call Artificial Texts. I argue that the only theory of authorial intent that can explain how Artificial Texts acquire literary meaning is Postulated Author Hypothetical Intentionalism.

Disclosure of Sources of Funding and Conflicts of Interest

This research was funded in part by a Social Sciences and Humanities Research Council grant.

The grant number is 767-2023-2658.

Statement of Originality and Disclosure of Use of Artificial Intelligence

In the preparation of this dissertation, no Artificial Intelligence tools were used.

Acknowledgements

This dissertation is dedicated to the memory of my father, Tom Vlaad, who always encouraged me to ask questions, and to my mother, Lise Vlaad, who continues to engage my work and ask me questions.

This work could not have been completed without the endless support and encouragement, scholarly and otherwise, from my community.

I'd like to thank my supervisor, Catherine Stinson, for their consistent critical engagement with this work as well as their support of my other scholarly pursuits. I'd also like to thank Elliot Paul, for being the best second reader a PhD candidate could ask for, and Marta Halina for serving as my host supervisor at the University of Cambridge in Spring 2024.

Further thanks go out to my fellow aesthetic philosophers, Stephanie Mills, and my artistic partner Steph Elms (GLITTERPUNK FOREVER)!

A special shout out to my sister, Marie Vlaad, and her partner, Jace Moore, and their lovely son Rudy (who is a cat).

Additional thanks go out to all of those who have engaged my work, supported me during hard times, shared in drinks and music, and told very good jokes: Adam Schipper, Alanna Veitch, Alisha Sharma, Andrew Lopez, Arthur Hill, Brennen Harwood, Colin McDermot, Dean Joseph, Isabel Xu, Jack Williams, Jaclyn Rekis, James Winslow, Jessica McMullin, Joel da Silva, Jordan Desmond, Justin Bouillon, Kaitie Jordeuil, Michael Luoma, Sadie Elliott, Yuanjin Xia, Ishaan Selby, Rowan Bell, Erik Nelson, Cassandra DeStellis, Jackie Davies, Lisa Guenther, Curtis Miller, Cody Rosevear, Shawn Levis, Sherri Elms, Christine Sypnowich, David Bakhurst, Rahul Kumar, and countless others.

Thanks to the many, many, group chats: Park Hangouts, The Bog, Album Club, The Fxllxs, Ape Theory, The Power of Friendship, Poetry Club, B Movie Night, and The Galaxy.

Thanks to the crews at The Grad Club and The Communist's Daughter.

Thanks to the Ethics and Technology Lab at Queen's University and the Periscope Lab at the University of Toronto.

Thanks to the support staff in the Philosophy Department at Queen's University: Jen McLaren, Sheena Wilkinson, and Roda Mendoza.

Thanks to the members and organizers of CSWIP, CSTP, and TPP.

A very special thank you to my partner, Ember Alvo, for their continued love and encouragement.

Finally, none of this would have been possible had Shannon Dea not supported my interest in philosophy as an undergraduate student at the University of Waterloo many years ago, for which I will be forever grateful.

Table of Contents

Abstract.....	i
Disclosure of Sources of Funding and Conflicts of Interest	ii
Statement of Originality and Disclosure of Use of Artificial Intelligence	iii
Acknowledgements.....	iv
List of Figures.....	ix
List of Abbreviations.....	x
Chapter 1: Art, Authorship, and Artificial Intelligence	1
1.1. The Problem.....	1
1.2. The Argument(s)	5
1.3. The Outline	7
Chapter 2: A Portrait of the Artist as a Young Algorithm.....	9
2.1. Can AI-Generated Images Sometimes be Considered Art?	9
2.2. AI Image Generators	11
2.3. Three Accounts of Art.....	16
2.3.1. Exhibited Features Accounts	17
2.3.2. Genetic Features Accounts.....	19
2.3.3. The Buck Passing Account	20
2.4. Understanding <i>The Electrician</i> as Art.....	22
2.4.1. <i>The Electrician</i> and Exhibited Features Accounts	23
2.4.2. <i>The Electrician</i> and Genetic Features Accounts	25
2.4.3. <i>The Electrician</i> and The Buck Passing Account.....	26
2.5. Who is the Artist?.....	27
2.5.1. Theories of Authorship	28
2.5.2. Conjunctive Authorship	29
2.5.3. DALL-E 2 as Computer Art.....	31
2.6. Implications.....	33
Chapter 3: AI Art is (Mostly) Bad.....	35
3.1. An Antinomy of Artificial Art.....	35
3.2. Mass AI-Art	35
3.2.1. The Nature Argument	36
3.2.2. The Intention Argument.....	37
3.3. Counter Arguments	38

3.3.1. Rejecting the Nature Argument	38
3.3.2. Rejecting the Intention Argument.....	41
3.4. Bad Art.....	43
3.4.1. The Failed Intentions View	44
3.4.2. The Failed Organic Whole View	46
3.5. Mass AI-Art as Bad Art.....	47
Chapter 4: Misalignment.....	50
4.1. The Aesthetics of AI Art.....	50
4.2. Weirdness	53
4.2.1. Weirdness as Non-Human Failure	53
4.2.2. Weirdness as Misalignment	55
4.3. Bias-Prompting	56
4.3.1. Algorithmic Bias	57
4.3.2. In/Visible.....	58
4.4. The Case for Misalignment.....	60
Chapter 5: AI, Medium, and Artform.....	63
5.1. Tool, Collaborator, or Something Else?	63
5.2. Artistic Medium	63
5.3. AI and Artistic Medium	66
5.3.1. Midjourney as Artistic Medium.....	66
5.3.2. Midjourney as Vehicle	67
5.4. AI Art and Medium Specificity.....	69
5.4.1. AI Art and MSV.....	70
5.4.2. AI Art and MSX.....	72
5.4.3. AI Art and MSF	73
5.5. Evaluating Helliwell’s Proposal.....	75
5.6. AI Art as Participatory Art	78
5.6.1. The Production Paradigm.....	78
5.6.2. The Exploration Paradigm	80
5.6.3. AI, Participatory Art, and Medium-Specificity.....	81
5.7. A Defence of AI Art	83
5.7.1. The Artificial Sublime	83
5.7.2. Misalignment	85

5.8. AI Art as Artform	86
Chapter 6: Texts Without Authors.....	88
6.1. Ascribing Literary Meaning in the Case of AI.....	88
6.2. Large Language Models.....	88
6.3. Three Kinds of Texts.....	90
6.4. Meaning and Artificial Texts.....	93
6.5. Six Views of Ascribing Literary Meaning	97
6.6. Fixing the Meaning of Artificial Texts.....	100
6.7. Consequences.....	102
Chapter 7: Towards a New Philosophy of AI Art.....	103
References.....	105
Artworks and Images Cited	123

List of Figures

Figure 1.1: Théâtre D'opéra Spatial (by Jason Allen)	1
Figure 1.2: Father Cat (by Alexander Mordvintsev).....	3
Figure 2.1: The Electrician (attributed to Boris Eldagsen and DALL-E 2)	9
Figure 2.2: The Diffusion Process (Taken from Yang et al. (2023) Figure 2).....	12
Figure 2.3: CLIP pre-training (Abbreviated Figure 1 from Radford et al. (2021)).....	14
Figure 2.4: unCLIP (taken from (Ramesh et al. 2022) Figure 2).....	15
Figure 3.1: 9/11 gender reveal (Creator Unknown)	39
Figure 3.2: Six Grandfathers.....	40
Figure 4.1: Blonde Braids Study II (by Minnie Atairu).....	59

List of Abbreviations

AI	Artificial Intelligence
LLM	Large Language Model
GAN	Generative Adversarial Network
TTI	Text-to-Image
DDPM	Denosing Diffusion Probabilistic Model
CLIP	Contrastive Language-Image Pre-Training
G-Art	Generative Art
CG-Art	Computer Generative Art
D12	12-Sided Die
A, B	Author
w	Work
F	Work-Description
MSV	Medium-Specific Evaluation
MSX	Medium-Specific Explanation
MSF	Medium-Specific Artform
EAI	Extreme Actual Intentionalism
MAI1	Modest Actual Intentionalism 1
MAI2	Modest Actual Intentionalism 2
AAHI	Actual Author Hypothetical Intentionalism
PAHI	Postulated Author Hypothetical Intentionalism

Chapter 1: Art, Authorship, and Artificial Intelligence

1.1. The Problem



Figure 1.1: Théâtre D'opéra Spatial (by Jason Allen)

At the 2022 Colorado State Fair, an artist named Jason Allen was awarded a prize for his artwork *Théâtre D'opéra Spatial* (Figure 1.1). It was later discovered that Allen had used Artificial Intelligence (AI) to create the image. This led to lots of backlash, with many claiming that Allen did not deserve the award as he did not create the image himself. Allen used a generative AI model called Midjourney to create the image, and many people saw this as an illegitimate way to create art. The news of Allen's story marked a shift in public awareness of AI tools, as more tech companies began to develop their own AI image generators. The internet became flooded with both a myriad of AI-generated images and an increasing amount of think pieces about what some had begun to call *AI art*.

AI is not new—computer scientists have been working on and with variations of the technology since the 1950s. However, the latest crop of advancements in AI—coupled with the ease of access provided by the internet—has led to many more people being able to use far more effective technology than at any prior point in the history of AI. AI is ubiquitous. It has applications in precision medicine, voice assistants such as Alexa or Siri, and every undergraduate student at university is well aware of the temptation to use ChatGPT to write their essay in time for the deadline. From Large Language Models (LLMs) to Generative Adversarial Networks (GANs), AI is *everywhere*.

This wasn't always the case. When I wrote the proposal for the grant that would end up funding this research, ChatGPT did not exist, and AI image generators still had not mastered the art of generating legible faces—or anatomically appropriate human hands. Much has changed in the past 3 years, and AI is both more advanced and more omnipresent than it was even in 2022. As more time passes, we might come to see Jason Allen's victory as a major turning point in public AI discourse. Or not. Whether the current AI bubble bursts seems incidental to the effect that the technology has already had on popular culture.

There are many different kinds of works that have been ascribed to the category of AI Art. Contemporary AI Art dates back at least as far as 2015 and the advent of DeepDream. Developed by Google engineer Alexander Mordvintsev, DeepDream is a computer vision program that when fed an image enhances textures and elements of the image to create a new image that is more dreamlike and psychedelic in quality. The developers refer to this technique as *Inceptionism* (Mordvintsev et al. 2015). While many images have been created using this method, *Father Cat* (Figure 1.2) remains a notable early example.

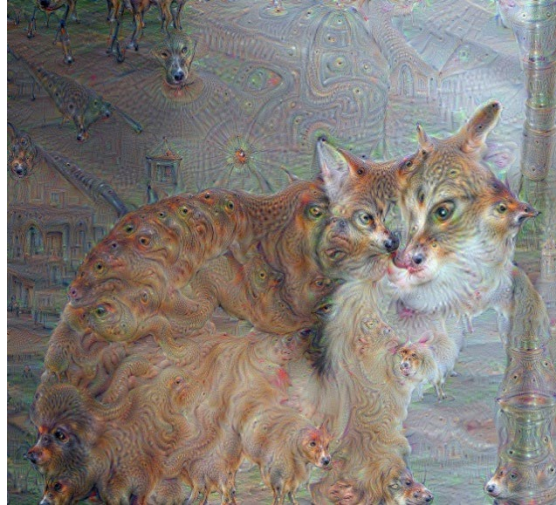


Figure 1.2: *Father Cat* (by Alexander Mordvintsev)

Elsewhere, Anna Ridler’s *Mosaic Virus* is a work of conceptual AI art made with the use of GANs. The artwork consists of AI-generated tulips that become discoloured with disease when the market value of bitcoin increases. Meant to evoke the 17th Dutch tulip commodity bubble—that formed when many of the tulips began contracting a virus that left them with aesthetically interesting discolorations that led them to being perceived as more valuable—*Mosaic Virus* brings historical market speculation into conversation with contemporary capitalist experiments such as cryptocurrency.

Another kind of AI Art comes in the form of literary texts produced by LLMs. We might consider works like David Jhave Johnston’s *ReRites*, a collection of AI-generated poetry. *ReRites* is a twelve-volume poetry set compiled between 2017 and 2018 consisting of AI-generated output arranged and edited by Johnston and later released in an edited format alongside critical essays by Johnston (2019). Allison Parrish’s *Articulations* (2018) takes a similar approach by training an LLM on a dataset of public-domain poetry and publishing a curated set of results. Works such as these—and electronic literature more broadly—have been taken up in academic

circles among literature scholars as both subjects of criticism (MacBeth 2023) and texts that challenge traditional understandings of literature (Hayles 2018). These are but a few examples of the ways the term that AI Art has been applied.

These examples of AI art open up at least three sets of questions. First, there is a set of aesthetic questions. Popular science fiction author Ted Chiang (2024) has famously argued that current AI models cannot create art. While there are scholars who agree with Chiang, such as Cascales (2023) and Nannicelli (2025), others (Wojtkiewicz 2023; Vlaad 2024) are more affirming of AI-generated images as art. Nonetheless, it remains an open question as to whether AI-generated images can be considered artworks proper. Elsewhere scholars have argued over whether AI image generation constitutes a new artistic medium (Young and Terrone 2025; Helliwell 2025), or whether we can understand it under existing artforms (Cross 2024). If AI-generated images are artworks, then to which of the arts do they belong? Finally, we might wonder whether AI-generated images have unique aesthetic values. Both Helliwell (2023) and Rini (2024) argue that they do.

Second, there is a set of questions concerning how we ought to adjudicate authorship in the case of AI-generated works (both text-based and image-based). While scholars are hesitant to assign author status to AI itself (Charles 2025; Schmidt 2025), they are also hesitant to grant such status to AI users (J. Moore 2025). Some have suggested forms of collaborative authorship (Khosrowi et al. 2025; Anscomb 2025), while others still have argued that AI is instead a bizarre case wherein there are no relevant authors of which to speak (Vlaad 2025). AI opens up new questions with respect to the philosophy of authorship and reignites old debates.

Finally, there are ethical questions concerning the production of AI art. We might wonder whether AI art constitutes forgery (Helliwell 2023), plagiarism (Marcus and Southen 2024), or theft (Goetze 2024). In particular, Goetze (2024) has argued that AI art constitutes creative labour theft. In this account artists mix their creative labour with creative building blocks (such as language, concepts, ideas, etc.) to make creative products. AI art is predicated on the theft of this creative labour through its unauthorized use of artists works as training data.

We might also have concerns about the outputs of AI artmaking tools (re)producing harmful biases. Scholars refer to this phenomenon as *algorithmic bias* (Danks and London 2017; Fazelpour and Danks 2021; Stinson 2022). Bias in training set data leads to biased outputs. If algorithmic bias cannot be mitigated, then AI art will necessarily be biased in some way(s). This could manifest as the generation of racist or sexist images, or a lack of social diversity represented in the outputs of AI. I explore AI artmaking and algorithmic bias in more depth in chapter four.

Furthermore, there are environmental concerns. AI requires a lot of energy and resources to operate. In the era of climate catastrophe it is reasonable to be concerned with technology that might lead to greater pollution and carbon emissions. Despite all of these concerns, scholars such as Ivanova (2025) have made the case that AI art might still be a worthwhile endeavour should we address these ethical issues.

1.2. The Argument(s)

This thesis is firmly rooted within the intellectual tradition of analytic philosophy. While scholars in other disciplines—such as cultural studies and media studies—have also explored the topic of AI art, I see myself in conversation with philosophers such as Helliwell (2023, 2025),

Wojtkiewicz (2023), and Cross (2024), who all approach questions of AI art with the tools of analytic philosophy.

This thesis makes a series of interrelated arguments. I begin by exploring the question of whether AI-generated images can count sometimes as artworks, and how we ought to adjudicate authorship in such cases. I argue that AI-generated images can and do sometimes count as artworks and suggest some author candidates.

I continue exploring whether AI-generated images can be artworks by considering a counter argument in the form of Nannicelli (2025), who holds that our aesthetic engagements with what he calls mass AI-art are best understood as similar to our aesthetic engagements with nature. As such, Nannicelli holds that mass AI-art is not properly art. I argue contra Nannicelli that such works are better conceived as instances of *bad art* rather than instances of *non-art*.

I continue by considering whether AI art possesses any unique aesthetic values. I argue that there exists a unique style of AI art that I term bias-prompting. This style exploits the algorithmic bias (Danks and London 2017; Fazelpour and Danks 2021; Stinson 2022) present in AI to highlight how such models can (re)produce harmful racist and sexist stereotypes. I make the case that bias-prompting is capable of producing a unique aesthetic feature I call misalignment.

I then engage Helliwell's (2025) argument that AI constitutes a unique artistic medium as well as an artform. Drawing on Davies (2009) I argue that AI is the vehicle through which the artistic medium of AI art is expressed. I defend a modified version of Helliwell's thesis against an alternative proposed by Cross (2024) that conceives of AI art as a kind of participatory art.

Finally, I examine the question of how to adjudicate authorship in the case of AI-generated texts. I draw a tripartite distinction between what I call Illusory Texts, Traditional Texts, and Artificial Texts. I argue that AI-generated Texts belong to the third group and are distinguished by their lack of an author. Rather than be created by an author, I contend that such texts are discovered by a curator. I then argue the only way to account for how these texts might acquire meaning is to defer to a hypothetical author collectively constructed by an audience of readers.

The dissertation concludes with a summary of the view and a consideration of further research questions.

1.3. The Outline

This dissertation is comprised of seven chapters, including this introduction and a conclusion. What follows are brief descriptions of the subsequent chapters that make up this project.

Chapter Two serves as a preliminary investigation into the major questions that govern my investigation, and in some ways, it serves as a microcosm of the dissertation as a whole. Can AI-generated images be artworks? Yes.

Chapter Three presents Nannicelli's view of mass AI-art as non-art. I argue that this is the wrong way to understand mass AI-art. Instead, I propose that we understand the phenomenon as an instance of bad art.

Chapter Four explores the aesthetics of AI art. I argue that some AI art possesses a unique aesthetic feature that I call generative anger. We can find this feature in a style of AI art that I call bias-prompting.

Chapter Five presents Helliwell's view of AI as a new artistic medium. I argue that this is the wrong way to understand AI. Instead, I propose that we understand AI as the vehicle through which the artistic medium of and artform of AI art is expressed.

Chapter Six shifts the focus to LLMs. Who is the author of AI-generated Texts? No one, but the prompt-giver is the curator. Which account of ascribing literary meaning can explain how AI-generated Texts acquire meaning? Postulated Author Hypothetical Intentionalism.

Finally, the dissertation concludes with a summary of my view and a gesture towards future directions of research.

Chapter 2: A Portrait of the Artist as a Young Algorithm

2.1. Can AI-Generated Images Sometimes be Considered Art?



Figure 2.1: The Electrician (attributed to Boris Eldagsen and DALL-E 2)

Something shocking happened at the 2023 Sony World Photography Awards. The winning submission was from a photographer named Boris Eldagsen (Parshall 2023). Titled *The*

Electrician (Figure 2.1), the photograph depicts two women in black and white. However, Eldagsen did not take the photograph himself. In fact, *The Electrician* is not a photograph at all. Rather, Eldagsen used an AI platform called DALL-E 2 to generate the image. Eldagsen refused the award, stating that the submission was a stunt to see if the competition could identify AI-generated images. Clearly, it did not. Eldagsen used his victory as a platform to start a discussion of the role of AI in photography—and art more broadly. The story made international headlines and public response was mixed. While some applaud and encourage the use of AI technologies in the creation of art, others question whether such creations can even be considered art at all. Eldagsen himself championed the use of AI to generate images akin to photographs as a new art form dubbed promptography.

The response to AI image generators has not been all positive however, with many taking to social media to express their disdain for the medium and to dismiss such images as non-art (Shaffi 2023). As far back as 2019, Sean Dorrance Kelly, a philosophy professor at Harvard wrote an article for the *MIT Technology Review* arguing that AI cannot be considered an artist, as it lacks the creative abilities found in humans (Kelly 2019). Contrary to Eldagsen, many people either dismiss AI-generated images as non-art or dismiss DALL-E 2 as an artistic collaborator.

At present, AI image generators are ubiquitous. From Midjourney to DALL-E 2, one need only log on to social media to become inundated with images created with generative AI. As the case of *The Electrician* shows, while some people believe that at least some AI-generated images ought to be considered art, others remain unconvinced. Additionally, there do not seem to be any straightforward ways of adjudicating which AI-generated images—if any—we ought to understand as art. We might expect to find some solutions by turning to philosophical definitions of art. However, in this chapter I will argue that while current trends in philosophy of art have

little difficulty understanding AI-generated images as art, there remains strong resistance from the general public accepting them as such. In order to explain this hesitancy, I suggest that the case of AI-generated images—such as *The Electrician*—presents us with difficult questions about *authorship* in ways that deeply challenge our intuitions of the concept.¹

This chapter proceeds as follows. First, I give a brief explanation of how AI image generators like DALL-E 2 work. Next, following Dominic Lopes (2014), I identify three kinds of theories of art: exhibited features accounts, genetic features accounts, and Lopes' buck passing account. I then show that all of these prevailing theories can satisfactorily deal with the case of so-called AI art by applying them to *The Electrician*. I argue that despite the fact that our current accounts of art all classify *The Electrician* as a work of art, resistance to this claim speaks to important issues that arise about the locus of authorship in such cases that might make us question their status as art. I outline the issues that arise when we consider any single individual to be the author in these cases, and present two complementary solutions. First, we can conceive of AI-generated images as collaborative works of art in a conjunctive sense. Additionally, we can consider AI image generators themselves to be artworks authored by the developers. I outline both views and conclude by exploring some implications that follow.

2.2. AI Image Generators

Contemporary AI image generators are a subset of deep generative models called vision-language models (Yang et al. 2023). As the name suggests, a popular application of these models is Text-to-Image generation, wherein an image is generated from a corresponding descriptive text (Du et al. 2022). A user inputs text prompts that the model then translates into images. These models are trained on datasets consisting of millions or even billions of image-text pairs

¹ Professor Kelly's (2019) arguments against AI as artists serves as a salient example of this phenomenon.

(Schuhmann et al. 2022). For example, one might have an image of a dog with the corresponding text “dog.” These image-text pairs help train the model to associate input text with sufficiently relevant corresponding images. Typically, models use an encoder-decoder framework. First, an encoder reads the data and transforms it into a different representation. Then, a decoder transforms the new representation into output (Ridoy et al. 2024). An example of this would be an encoder transforming image-text pairs (input) into vectors (representation) which are then transformed into new images (output) by the decoder.

State of the art image generation often uses diffusion models (Ho et al. 2020; Yang et al. 2023). Introduced by Sohl-Dickstein et al. (2015) and improved upon by others (Ho et al. 2020), Denoising Diffusion Probabilistic Models (DDPMs) work by introducing signal noise into the data and slowly reversing the process to generate new data samples. In the case of Text-to-Image generation we might imagine the text prompt “dog” gets matched with a corresponding image of a dog from the dataset. The image-text pair is then transformed into a vector by an encoder. The diffusion model then introduces noise interference. The interference is slowly reversed by a decoder that then transforms the vector into a new image reconstructed based on the likelihood of certain elements being closer in proximity to others (see Figure 2.2).

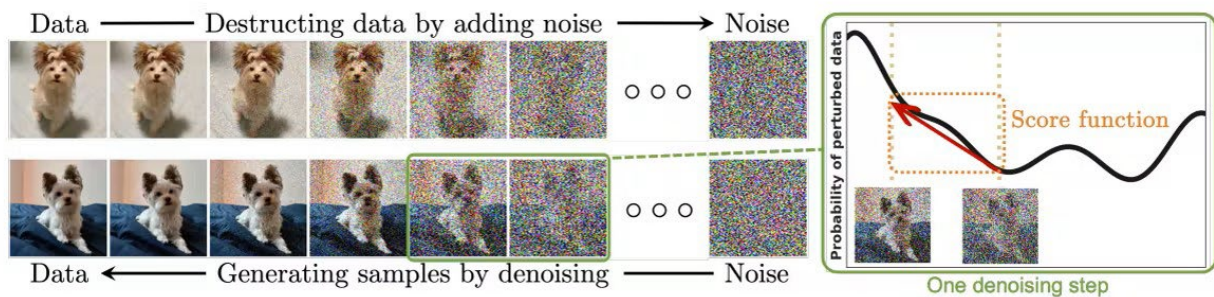


Figure 2.2: The Diffusion Process (Taken from Yang et al. (2023) Figure 2)

For example, the brown pixel of a dog's ear has a low probability of being located next to a white pixel of fur from the dog's tail, so the model is unlikely to generate an image where this is the case. However, there is a high probability that two brown ear pixels will be next to each other and thus it is highly likely that the model will generate an image concordant with this high probability. Diffusion models tend to achieve better results over time and as such are used in many vision-language models—including DALL-E 2.

Elsewhere, Radford et al. (2021) have proposed a vision-language model called CLIP (Contrastive Language-Image Pre-training) designed to classify images by giving corresponding text. Originally conceived as a tool for automated image captioning, CLIP uses both an image encoder and a text encoder. During pre-training, image-text pairs are fed to CLIP. The image encoder translates the images into vector representations which are then arranged in a $1 \times N$ matrix. The text encoder does the same for the text, arranging the vectors in an $N \times 1$ matrix. The two matrices are multiplied to generate an $N \times N$ vector space (see Figure 2.3). The elements that comprise the space are image-text pairs. Image-text pairs that are similar to one another are closer to each other in the vector space. When fed an image I_n , CLIP matches it with the closest image-text pair $I_n T_n$ and produces the corresponding text T_n as output.

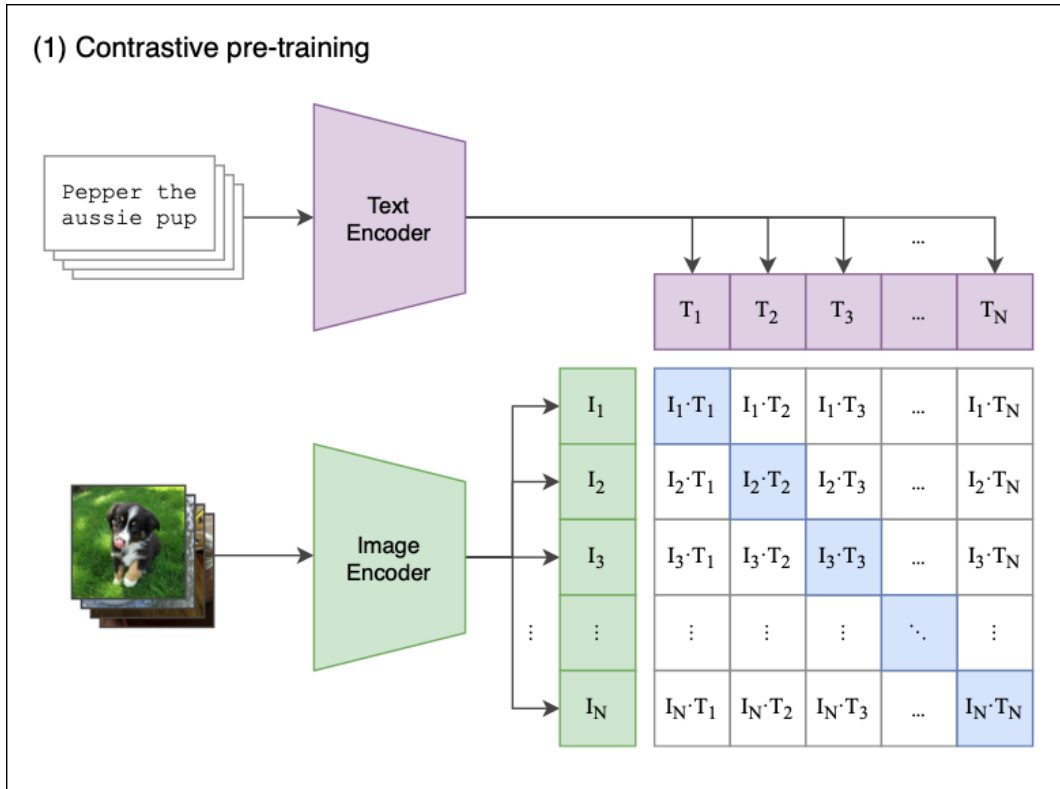


Figure 2.3: CLIP pre-training (Abbreviated Figure 1 from Radford et al. (2021))

For example, we might have many image-text pairs of dogs, but only some of those dogs are running. An image of a running dog is closer to the image-text pair of a running dog with the caption “running dog” than it would be to an image-text pair of a sitting dog with the caption “sitting dog.” Thus, when fed an image of a running dog, CLIP generates a caption along the lines of “running dog.”

Ramesh et al. (2022) have suggested a two-step image generation process involving a prior model that produces CLIP-based image embeddings and a diffusion decoder that transforms the embeddings into images. This technique is called unCLIP—as it inverts the CLIP process—and is used by DALL-E 2. In step one, a text prompt is transformed into an associated CLIP-based image embedding by the prior model. For example, the prompt “running dog” will

generate embeddings based on training set images for which “running dog” is an appropriate caption. In step two, the diffusion decoder transforms the image embeddings into new images. The user can then choose from a set of images produced which one they like best, or they can refine their prompt to produce a different set of images. Whereas CLIP generates text from image inputs, unCLIP generates images from text inputs. Figure 2.4 illustrates the training process above the dotted line and the generation process below the dotted line.

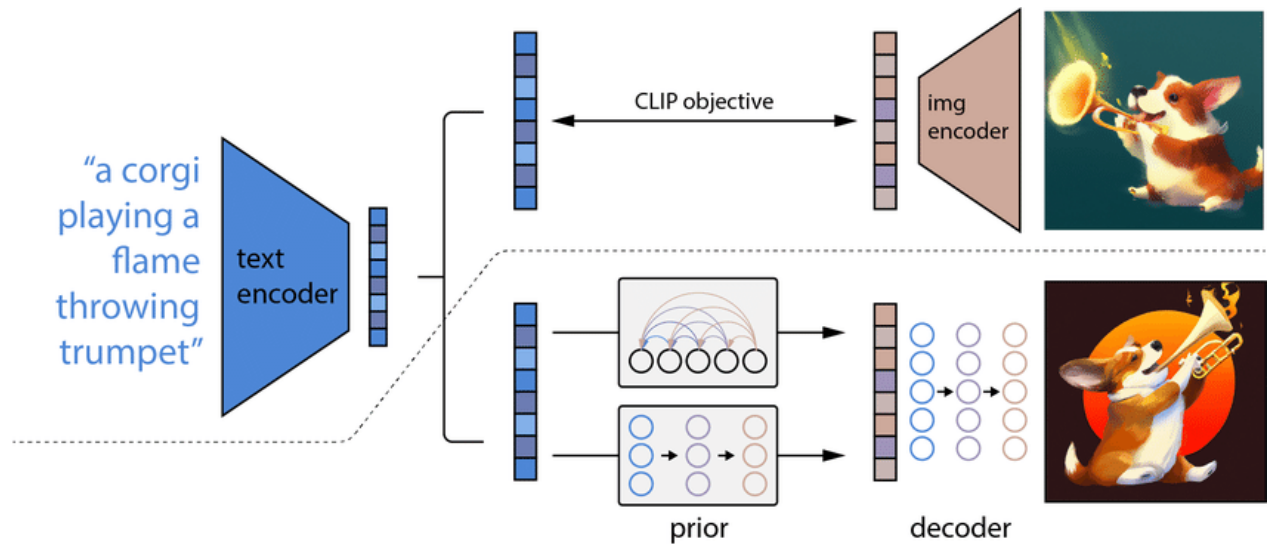


Figure 2.4: unCLIP (taken from (Ramesh et al. 2022) Figure 2)

While DALL-E 2 has the capacity to generate fascinating images such as *The Electrician*, it remains an open question as to whether some of these images ought to be considered art. In the next section I give an overview of three leading contemporary accounts of art to guide this discussion.

2.3. Three Accounts of Art

Philosophers and art critics have long debated what constitutes art. For Kant ([1790] 2009), fine art is the product of genius. It is both novel and exemplary of the form. The Beauty of art thus differs from the Beauty of nature as it has no final end but itself. Some thinkers (Meynell 1986) claim that aesthetic value in the form of Beauty is a necessary property of good art. This view has been contested (Steiner 2001; Danto 2003), with scholars arguing both that we ought to consider antiaesthetics (Ziff 1984) and negative aesthetic value (Brady 2011). The former holds that there is art without any aesthetic value present at all, whereas the latter holds that some aesthetic value is negative in quality. Rather than judge a work of art as exemplary in virtue of its Beauty, we might judge it as exemplary in virtue of its Ugliness. Others still reject the concept of art entirely (Richter 1965), a move common among practitioners of Dada. Recently some scholars have attempted to reclaim Beauty (Nehamas 2007; Lopes 2018), but the conversation has largely shifted to defining art in ways other than the presence or absence of aesthetic value.

In *Beyond Art* (2014), Dominic Lopes addresses a problem that arises in contemporary trends in analytic philosophy of art.² Lopes takes there to be two main kinds of theory of art. On the one hand, there are exhibited features accounts, that hold that we can identify art by its properties. On the other hand, there are genetic features accounts, that hold that we can identify art by the context of its genesis. As there are paradigm examples of art that either exhibited features or genetic features accounts fail to classify as art, neither account can provide a

² There are other ways of carving out current trends in analytic philosophy of art, such as the inclusion of hybrid exhibited/genetic features accounts as distinct kinds of accounts. See Longworth and Scarantino (2010) for an example of such an account. However, given that these accounts are ultimately comprised of exhibited and genetic features accounts as Lopes describes them, and given the uniqueness of Lopes' own account of art, it is reasonable to follow Lopes as other views are both reducible to Lopes' taxonomy, and would have to discuss the buck passing account regardless.

comprehensive understanding of art as such. Lopes proposes the buck passing account in order to solve this problem. While the buck passing account is able to solve the problem of classifying art, it does not solve the demarcation problem of art—nor does Lopes claim it can do so.

I explore each theory in depth, beginning with exhibited features accounts, followed by genetic features accounts, and concluding with the buck passing account.

2.3.1. Exhibited Features Accounts

Exhibited features accounts define art as possessing a set of properties. While these properties are often aesthetic in nature, that need not necessarily be the case (Lopes 2014). For example, Paul Ziff (1953), proposes seven sufficient properties of a work of art:

1. It is recognizable as a work of art.
2. It was made by an artist deliberately with skill and care.
3. It was intended by the artist to be treated like a work of art.
4. It is presented to others by whom it is discussed, studied, contemplated, admired, and criticized.
5. It contains subject matter.
6. It has a complex formal structure.
7. It is good of its kind. (Ziff 1953, 60–61)

It is important to note that in Ziff's account a work of art need not possess all seven properties. If a work were to possess all seven, then we know for certain that it is a work of art. Accordingly, if a work does not exhibit any of the features listed, then it is not a work of art. However, if a work exhibits some of the features in question, then we might make a case for the work as art.

Consider Van Gogh's *The Starry Night* as an example of a work that meets all seven criteria:

1. *The Starry Night* is recognizable as a painting.
2. It was made by Van Gogh deliberately with skill and care.
3. It was intended by Van Gogh to be treated like a work of art.
4. *The Starry Night* has been presented to others by whom it has been discussed, studied, contemplated, admired, and criticized for over a century.
5. *The Starry Night* contains subject matter—it depicts a night sky in Paris.
6. *The Starry Night* has a complex formal structure—Van Gogh used a myriad of impressionist techniques in its creation.
7. *The Starry Night* is a good painting.

As *The Starry Night* is a paradigm case of art, it is unsurprising that it meets all seven of Ziff's criteria. While exhibited features accounts such as Ziff's provide useful heuristics for identifying many—if not most—works of art, they struggle to deal with certain difficult cases such as Marcel Duchamp's *Fountain*.

In 1917 Duchamp took a readymade sculpture of a urinal and entered it in an art gallery under a pseudonym. This caused a great deal of controversy and sparked debates about whether *Fountain* ought to be considered art at all, as it was merely an average urinal presented as a provocation. According to exhibited features accounts such as Ziff's, *Fountain* does not seem to qualify for art status, as it meets few of the requisite criteria. However, contemporary scholars *do* consider *Fountain* to be a work of art, perhaps being an early example of what would later come to be known as conceptual art. That exhibited features accounts fail to classify avant-garde works such as *Fountain* as art seems to be a shortcoming of such accounts.

2.3.2. Genetic Features Accounts

Genetic features accounts, however, do not share this shortcoming. While exhibited features accounts focus on properties expressed by a work, genetic features accounts define art in terms of the context of its creation. It is to these accounts that I now turn.

Genetic features accounts come in two variations. The first are institutional accounts. These hold that for a work of art to be considered as such, it must be presented to an artworld (Danto 1964; George Dickie 1974). This means that the work in question must be presented to an audience of artists, critics, and aficionados. For a work to be a work of art, it must have an intentional connection to art as an institution. Note that this does not mean that the artist or creator of the work must intentionally understand it as art. Rather, anyone can take the work and present it to an artworld.

Dickie (1969) himself gives an example in the form of works made by chimpanzees. According to Dickie, if chimpanzees were to make paintings—an event which has been known to occur on occasion—then what matters is what is done with the paintings. If the paintings were to be displayed in the study of a primatologist, then we ought not consider them art. Rather we might conceive of the paintings as artifacts providing a glimpse into the cognitive and motor capacities of chimpanzees. However, if the paintings were instead displayed in an art gallery, and thus presented to an artworld, then they would have a claim to art status and likely be understood as works of art. We might similarly think of a state of affairs wherein *The Starry Night* remained in Van Gogh's attic and was never presented to an artworld. If that were the case, then *The Starry Night* would not count as art according to this account. Levinson (1979) presents this as a reductio of Dickie's account, and develops an historical account of art in response.

The second variation of genetic features accounts are historical accounts. These hold that for a work of art to be considered as such it must stand in relation to a history of art (Levinson 1990). For example, a Modernist poem might be responding to a tradition of Romantic poetry, thus standing in historical relation to earlier works of art. Tom Stoppard's play *Rosencrantz and Guildenstern are Dead* is an example of a more explicit response to histories of art, as it is a retelling of Shakespeare's *Hamlet* from the perspective of the minor characters Rosencrantz and Guildenstern.

An upshot of both variations of genetic features accounts is that they can classify cases of avant-garde works as art. Whereas exhibited features accounts fail to understand *Fountain* as a work of art, for example, genetic features have no such issue. *Fountain* was both presented to an artworld when it was displayed in a gallery, and it exists in relation to a history of avant-garde and experimental art. However, exhibited features accounts would not struggle with the case of *The Starry Night in the attic*, as it would meet most of the sufficient criteria. It is especially important that it meets the first of Ziff's criteria in that it is recognizable as a painting. If the inability of exhibited features accounts to classify works such as *Fountain* as art counts as a weakness of these accounts, then the inability of genetic features accounts to classify *The Starry Night in the attic* as art would seem to be a weakness of those accounts as well.

While neither exhibited features nor genetic features accounts can easily deal with challenging cases, Lopes provides a solution in the form of the buck passing account, to which I now turn.

2.3.3. The Buck Passing Account

In *Beyond Art* (2014) Lopes provides an alternative to both exhibited features and genetic features accounts of art. According to Lopes, the buck passing theory of art "passes the buck"

from a general theory of art to specific theories of the arts. In this account, “x is a work of art = x is a work of K, where K is an art” (Lopes 2014, 49). “The Wasteland” by T. S. Eliot is a work of poetry, where poetry is one of the arts. Therefore, “The Wasteland” is a work of art. The buck passing account of art displaces the burden of explanation from a general theory of art to specific theories of the arts.

Lopes defends the buck passing account of art against both exhibited features and genetic features accounts. According to Lopes, the buck passing account of art “...deals more effectively with the hard cases than its competition” (Lopes 2014, 46). By “hard cases” Lopes has in mind works such as *Fountain*. Artists and critics have since developed vocabulary to deal with the case of *Fountain*, with many—including Lopes—considering it to be a work of conceptual art.

Lopes argues that exhibited features accounts are forced to deny hard cases like *Fountain* as art, on the grounds that there will exist non-art counterparts to the hard cases that exhibit the same features. Put simply: other than the context in which we find them, there is no discernable difference in features between *Fountain* and any given urinal one might find in a public restroom. Genetic features accounts take the opposite position and accept *Fountain* as art. However, for Lopes, both genetic features accounts and exhibited features accounts—what he terms buck stopping accounts—fail to take the hard cases seriously. For Lopes,

buck stopping theories of art cannot cope effectively with the hard cases because the theories themselves and the criteria for choosing among the theories echo—and so cannot adjudicate—intuitions about the hard cases. (Lopes 2014, 63)

According to Lopes, exhibited features accounts do not take the hard cases seriously because they unilaterally dismiss them as non-art. This dismissal runs contrary to how works such as

Fountain are understood by artists and critics themselves. On the other hand, genetic features accounts accept the hard cases as art a priori, leaving no room for discussion.³

The foreclosure of contesting the status of the hard cases also runs contrary to how these works are understood. While *Fountain* is considered art by most professional artists today, this was not the case upon its unveiling in 1917. The buck passing account is appealing because it allows room for discussion and contestation, whereas the other types of accounts do not.

Additionally, the buck passing account can also deal with the case of the chimpanzee paintings stored in the primatologist's study. Applied to the buck passing account, the chimpanzee paintings are works of painting, where painting is one of the arts. Therefore, the chimpanzee paintings are works of art. Likewise for *The Starry Night* that has been left in the attic. The fact that the buck passing account can classify challenging cases that either exhibited features or genetic features accounts struggle with is a definite upshot of the account.

However, there is a downside to the buck passing account. Namely, in passing the buck to theories of the arts, it does nothing to solve the demarcation problem of art. Furthermore, if the case in question is not currently legible as one of the arts, then the buck passing account has nothing to say about it. It is easy to classify works like *Fountain* as conceptual art when one has the vocabulary and framework available to do so. It is another situation entirely where no such vocabulary or frameworks presently exist. While the buck passing account allows room for contestation and discussion, it cannot meaningfully contribute to said discussion if the relevant artform has not yet been articulated.

2.4. Understanding *The Electrician* as Art

³ There are many examples of contested cases of demarcation problems in the arts, such as photography (Scruton 1981; Lopes 2016) and videogames (Smuts 2005; Ensslin 2012). After developing the buck passing account in *Beyond Art* (2014), Lopes gives an analysis of videogames using his account.

I now show how the contemporary accounts of art discussed in this chapter come to understand *The Electrician* as a work of art. I begin with exhibited features accounts (exemplified by Ziff), then turn to genetic features accounts (both institutional and historical) and conclude with Lopes' buck passing account. I argue that each kind of account classifies *The Electrician* as art—albeit for different reasons. This section should be understood as a small case study rather than a comprehensive survey on how all accounts of art would understand *The Electrician*.

2.4.1. *The Electrician* and Exhibited Features Accounts

I begin with exhibited features accounts. Recall Ziff's (1953) seven sufficient properties for whether a work qualifies for art status:

1. It is recognizable as a work of art.
2. It was made by an artist deliberately with skill and care.
3. It was intended by the artist to be treated like a work of art.
4. It is presented to others by whom it is discussed, studied, contemplated, admired, and criticized.
5. It contains subject matter.
6. It has a complex formal structure.
7. It is good of its kind. (Ziff 1953, 60–61)

It seems to be the case that *The Electrician* meets most of these conditions. Taken one at a time, we see that:

1. *The Electrician* is clearly recognizable as a work of art, insofar as it is recognizable as a photograph. It need not actually *be* a photograph for this criteria to obtain. Note that it need only be *recognizable* as such.

2. It is unclear whether we can say that *The Electrician* was made by an artist with skill and care, given that we would first have to determine to whom we ought to attribute the piece. The work is credited to both Eldagsen and DALL-E 2. It seems strange to attribute the qualities of skill and care to AI, and it is also unclear whether the prompt curation done by Eldagsen counts as such.
3. Eldagsen intended for *The Electrician* to start a conversation about art, serving a role similar to *Fountain*.
4. *The Electrician* was submitted to the Sony World Photography Awards, where it was discussed, studied, contemplated, admired, and criticized by an audience of judges to whom it was presented.
5. Whether *The Electrician* contains subject matter depends on which account of depiction one endorses. Resemblance accounts such as those endorsed by Abell (2009) and Lopes (2004) would hold that *The Electrician* contains subject matter insofar as it depicts two women posing. Those who reject resemblance accounts of depiction, such as Goodman (1976), would deny that *The Electrician* contains subject matter.
6. Given the realistic quality and high resolution of *The Electrician*, it seems fair to say that it has a complex formal structure.
7. *The Electrician* was well received by the judges, winning the competition. This seems to be evidence that it is good of its kind (it is a good work of art).

Given that *The Electrician* clearly meets four out of the seven sufficient criteria provided by Ziff—with criteria 2, 3, and 5 being ambiguous or contentious—it seems safe to say that it has a reasonable claim to art status.

2.4.2. *The Electrician* and Genetic Features Accounts

Recall that genetic features accounts come in two varieties. First, there are institutional accounts that hold that a given work must be presented to an artworld for it to count as art. Second, there are historical accounts that hold that a given work must stand in a historical relation to art for it to count as such. I will argue that *The Electrician* counts as art according to both forms of genetic features accounts.

As *The Electrician* was submitted to the Sony World Photography Awards as a work of art, it was indeed presented to an artworld. In this case the artworld is comprised of judges, although the international news coverage that followed its victory arguably exposed it to an even larger artworld in the form of the general public.⁴

The Electrician also has ties to a history of art, as it seem to be an example of generative art (G-art) (Boden and Edmonds 2009). According to Boden and Edmonds, “G-art works are generated, at least in part, by some process that is not under the artist’s direct control” (Boden and Edmonds 2009, 37). This would include any kind of art generated through the use of generative systems. Aleatoric music—sometimes called chance music—comes to mind as a salient example. Composers such as John Cage would sometimes craft pieces of music through the use of chance operations. The paradigm example here is using dice as a tool in composition. Suppose you have a twelve-sided die (D12). Each number on the die corresponds to a note of the chromatic scale (1 = C, 2 = C[#], 3 = D, etc.), and rolling the die determines the next note in the sequence. This approach to artmaking is also seen in the generation of *The Electrician*. While DALL-E 2 is far more complex than a D12, both are forms of generative systems.

⁴ For an alternative perspective on how Danto would interpret AI art, see Cascales (2023).

2.4.3. *The Electrician* and The Buck Passing Account

In order for the buck passing account to classify *The Electrician* as a work of art, it must belong to one of the arts. In this section I argue that *The Electrician* is a work of computer generative art (CG-art), a subset of G-art. As such, *The Electrician* is a work of art.

In “What is Generative Art?” Margaret Boden and Ernest Edmonds (2009) identify a taxonomy of what they call generative art. While Boden and Edmonds identify eleven distinct but related types of generative art—sometimes called computer art—the relevant definition here is CG-art. According to Boden and Edmonds, “CG-art is produced by leaving a computer program to run by itself, with minimal or zero interference from a human being” (Boden and Edmonds 2009, 37). This definition certainly maps on to Eldagsen’s role in creating *The Electrician*. While Eldagsen provided the prompt that ultimately generated *The Electrician*, refining it along the way, it was DALL-E 2 that actually created the image.

One could argue that prompt curation is not constitutive of minimal interference. Rather, the very existence of *The Electrician* is dependent on Eldagsen’s prompts. Here we might turn to the analogy of building a house, where Eldagsen takes on a role equivalent to the homeowner whose house is being built. While both Eldagsen and the homeowner might guide the contractors and the AI by giving them suggestions—such as what color to paint the walls, or how many women to include in the image—it is ultimately the contractors and the AI that do the heavy lifting. If we would say that the homeowner in this case is minimally involved in building the house, then the same should hold of Eldagsen’s relationship with the creation of *The Electrician*.

In summary, *The Electrician* is a work of CG-art, where CG-art is one of the arts. Therefore, *The Electrician* is a work of art. Note that we could also say that *The Electrician* is a work of promptography if we so desire—bearing in mind that the term is new and contested.

According to all three kinds of accounts of art—exhibited features accounts, genetic features accounts, and the buck passing account—*The Electrician* ought to be considered a work of art. Note that it does not follow that *The Electrician* is a work of photography even if it is a work of art. However, there remains popular resistance to accepting works such as *The Electrician* and other AI-generated images as art. In the next section I suggest that we have unique reasons for being skeptical of AI-generated content as art given questions that arise about authorship in these cases.

2.5. Who is the Artist?

While the term author has diverse connotations within and outside the field of literature, in this context I am using the term to refer to creator of a given work. For this reason, I use the terms author and artist interchangeably. It is not readily apparent to whom we ought to attribute authorship in cases of AI-generated images. Recall that for Kant, the difference between the Beauty of fine art and the Beauty of nature is that the former is the product of genius. Regardless of how we interpret the term genius, it seems clear that Kant intends for it to refer to a person. If we cannot identify an artist this seems to be grounds for remaining skeptical of AI-generated images possessing art status, as it seems conceptually incoherent to have art in the absence of a corresponding artist. It is also relevant that one of the seven of the criteria proposed by Ziff that does not clearly obtain in the case of *The Electrician* concerns whether it was made by an artist with skill and care.

In this section I provide an overview of some accounts of authorship and then identify two possible solutions to the case of *The Electrician*. First, I consider that authorship is conjunctive in these cases. As such, both the developers and the prompt-giver are responsible for the art generated by DALL-E 2. Second, I consider that the images generated by DALL-E 2 are

not always themselves art. Rather, we can understand DALL-E 2 itself as a work of computer art (Lopes 2010). In this case, the images generated are the results of the audience—the prompt-givers—interacting with the artwork. I dismiss both of these solutions as unable to properly explain authorship in the case of AI. I conclude by gesturing towards a third possibility, that AI-generated works do not possess authors. I return to this idea in chapter 6.

2.5.1. Theories of Authorship

The philosophy of authorship has largely been concerned with ascribing author status in the case of literary texts, rather than visual artworks. In this section I give a brief overview of some of these accounts and how they might apply to the case of *The Electrician*.

Roland Barthes ([1967] 1977) famously argues that the author as such is dead. Literary works do not acquire their meanings from an all-knowing author-figure, rather, it is up to the audience or the readers to imbue a work with meaning. If we take the “author” to be the one who imbues a work with meaning, then perhaps the author is dead in the case of AI. However, this is not how we have been using the term. For our purposes, the “author” or “artist” is synonymous with the “creator” of a given work. Determining the locus of meaning making is somewhat beside the point in this case. For example, it is logically consistent that we might identify a “creator” (or “creators”) in the case of *The Electrician*, and this individual or group is not responsible for imbuing the work with meaning.

Elsewhere, for Foucault ([1969] 1998), all texts have writers, but not all writers are authors. Private correspondence is an example of texts with writers but not authors. Foucault conceives of an author-function that sets the terms of the discourse surrounding a given text. According to Foucault, this author-function must be abolished in order to transform the kinds of conversations we can have and questions we can ask of a text. On the other hand, Nehamas

(1986) argues against Foucault and holds that the author is co-constructed with a text and is themselves subject to interpretation. A consequence of this view is that the author cannot be abolished as they always emerge through the text.

Abolishing the notion of authorship might seem appealing in the case of *The Electrician*, as it would mitigate some of the hesitancy we have with attributing art status to the work. However, this once again seems to sidestep the issue of ascribing creator status. If we could abolish the author (or the author-function) we would still be left with a “writer” or “creator.” With this in mind, we can clearly see that the question we should be asking in the case of AI-generated art is not “who is the author” or “who is the artist” but “to whom do we attribute creator status?”

To this end I turn to work on collective authorship that takes the terms “author” and “creator” to be largely synonymous. We can add “artist” to that list as well.

2.5.2. Conjunctive Authorship

If we consider AI-generated images such as *The Electrician* to be works of art in their own right, attributing authorship to a single individual or class of individuals seems to be an untenable position. We cannot consider the prompt-giver to be the sole artist, as they neither generate the images nor design the system that does. In other cases of generative art, such as aleatoric music, the composer both designs and enacts the systems that generates the art. In the case of AI-generated images, the prompt-giver only enacts the system.

Likewise, we cannot consider the developers to be the sole artists, as the input of the prompt-giver is *necessary* to generate the artwork. For similar reasons, the model cannot be the sole author as it can only perform its functions thanks to human involvement—both from the

developers and the prompt-giver. As such, if we are to consider AI-generated images to be art, then a conjunctive view of authorship seems to be required.

The notion that authorship can be distributed amongst a group is not new. Marcel Duchamp ([1957] 1975) argues that all art is multi-authored, given that it is the audience that deciphers and interprets the art and thus contributing to its meaning. Barthes ([1967] 1977) gives a similar argument when he declares that the author is not the sole locus of meaning-making. Rather, the audience are active participants in the creation of a work: “the birth of the reader must be at the cost of the death of the author” (Barthes [1967] 1977, 148). This is not to say that the writer of a text plays no role, rather the act of authorship is a group endeavor.

We need not go as far as to say the audience co-authors every work of art. However, in the case of AI-generated images there are clear candidates for authorship such that we can identify the individual contributing artists. We might say that the AI-generated images are the work of at least two artists. These are the developers and the prompt-giver. We could possibly claim that the model also counts as a contributing artist. However, this would commit us to also assigning the role of contributing artist to the D12 in the case of aleatoric music. I remain agnostic with respect to the status of the model as artist.

Building on work on collective authorship in cases such as films, wherein works are co-created by multiple individuals (Stillinger 1991; Sellors 2007), Christy Mag Uidhir (2011) gives an account of authorship capable of accounting for attributing authorship in collaborations. What makes this account unique is that instead of considering authorship a two-place relation—an agent **a** is an author of a work **w**—Mag Uidhir takes authorship to be a three-place relation such that **a** is an author of **w** as a work-description **F**. For example, Charlie Kaufman is the author of *Being John Malkovich* as a script, and Spike Jonze is the author of *Being John Malkovich* as a

film. Mag Uidhir goes one step further and posits conjunctive authorship such that “Only [A and B...] is an author of *w* as an F, but neither A nor B... taken alone is an author of *w* as an F.” (Mag Uidhir 2011, 377). This account of conjunctive authorship maps neatly onto the case of *The Electrician*.

Given that neither Eldagsen nor the developers of DALL-E 2 are solely responsible for generating *The Electrician*, and yet both are necessary for its genesis, we can say the following. Only [Boris Eldagsen and the developers of DALL-E 2] is an author of *The Electrician* as a work of art, but neither Boris Eldagsen nor the developers of DALL-E 2 taken alone is an author of *The Electrician* as a work of art.

However, this framing elides the ways that artists who might contribute to the outputs of DALL-E 2 in virtue of their art featuring in the dataset. Given that the diffusion process creates novel images by drawing on existing ones, it is impossible to say which artworks and artists contributed to which particular pixels in a given AI-generated image. It is not only the case that a conjunctive account of authorship fails to include artists whose work is used as training data as contributors, the view is unable to identify which elements were contributed by whom.

2.5.3. DALL-E 2 as Computer Art

According to Lopes (2010), “an item is a computer art work just in case (1) it’s art, (2) it’s run on a computer, (3) it’s interactive, and (4) it’s interactive because it’s run on a computer” (Lopes 2010, 27). Let’s unpack this in relation to DALL-E 2. Condition (1) will be the most difficult to satisfy, so I will leave it for last. Condition (2) easily obtains—DALL-E 2 is run on a computer. Condition (3) also easily obtains—DALL-E 2 is interactive insofar as it requires text-prompts from its users. Condition (4) requires that DALL-E 2’s interactivity be necessarily linked to its being run on a computer. Given that DALL-E 2 is AI, which necessarily runs on a computer,

DALL-E 2 is interactive because it's run on a computer.. Ergo, Condition (4) obtains. With respect to Condition (1), we have pragmatic reasons for understanding DALL-E 2 as a work of (computer) art, as considering it as such changes how we understand both the images it generates and how we are to attribute authorship.

On the surface however, DALL-E 2 does not appear to be a work of art itself. We might appeal to the three accounts of art provided earlier to see how it fares under each. Exhibited features accounts present the most resistance to understanding DALL-E 2 as art. Some conditions clearly obtain. DALL-E 2 was made with skill and care, it has a complex formal structure, and it was presented to others by whom it was discussed, studied, contemplated, admired, and criticized. For some conditions it is more ambiguous if they obtain. It is unclear if DALL-E 2 is good of its kind, or what it would mean for it to have content. Furthermore, there is at least one condition that does not obtain at all—DALL-E 2 is not immediately recognizable as art. Even still, we can certainly say that DALL-E 2 has been embraced as a source of entertainment by many.

Genetic features accounts fare better—DALL-E 2 was presented to the public with an invitation to engage as artists. We might be able to interpret this as DALL-E 2 being presented to an artworld. Additionally, DALL-E 2 sits in a historical relationship to other kinds of computer art insofar as interacting with it as a prompt-giver yields outcomes similar to other forms of computer art. We might think of DALL-E 2 as being similar to some video games where the agency of the player determines—to an extent—how the world of the game develops. Lopes himself uses the example of video games in his discussion of computer art.

Finally, the buck passing account seems to have little difficulty understanding DALL-E 2 as art, as it can be classified under several different arts. We might say that DALL-E 2 is a work

of computer art, given that it seems to satisfy the conditions laid out by Lopes.. We might say that it is a work of conceptual art, with the concept being the democratization of art. Finally, we might say that it is interactive art—similar to performance art—given that it is an object whose form and content is affected by the behavior of the audience. All these ways of classifying DALL-E 2 render it a work of art according to the buck-passing account.

However, even if we take DALL-E 2 to be an artwork authored by the developers, this does not explain how individual outputs sometimes become artworks. Interactions with artworks by an audience are not typically understood as independent artworks. This seems to be the wrong approach, as works such as *The Electrician* do seem to meet the criteria for artwork status.

2.6. Implications

Given that AI-generated images such as *The Electrician* can be considered works of art, there is a strong case to be made that at least in principle, they ought to be allowed in art competitions. Furthermore, identifying what kind of art AI-generated images count as will determine what kinds of art competitions they ought to be allowed into. For example, since *The Electrician* is not a photograph, the Sony World Photography Awards is an inappropriate venue to which it could be submitted.

One solution would be to have separate competitions for AI-generated images as art. This would both provide a venue for sharing and appreciating such works as art without taking away from other forms of art such as photography. This would also discourage practitioners of AI art to disguise their works as paintings or photographs and might even encourage them to push the boundaries of AI-generated art into strange new directions unique to the medium.

However, the question of how we ought to adjudicate authorship remains. I return to this question in chapter 6 and suggest that AI-generated works possess curators rather than authors. In other words, AI-generated works are authorless. In the meantime, I turn my attention to another recent argument against AI-generated works as art.

Chapter 3: AI Art is (Mostly) Bad

3.1. An Antinomy of Artificial Art

Can AI-generated images be considered artworks? Some scholars say yes (Wojtkiewicz 2023; Rini 2024; Young and Terrone 2025), others say no (Cascales 2023; Nannicelli 2025). Public opinion is divided as well. Artists such as Eldagsen see AI as an exciting new innovation in artmaking (Parshall 2023), while science fiction author Ted Chiang argues that AI cannot make anything that matches human art (Chiang 2024). To help explain this disagreement, Moore (2025) articulates what he calls an *antinomy of artificial art*—a contradiction between three plausible positions:

1. AI-generated images can count as artworks.
2. All artworks have artists.
3. AI-generated images do not have a clear artist.

In this chapter I explore whether we can deny the first premise.

I begin by turning to Nannicelli’s (2025) work on *mass AI-art*, wherein he provides two arguments for denying the art status of AI-generated images. I call the first argument the *nature argument* and I call the second argument the *intentional argument*. I give reasons for rejecting each of these arguments and denying art status to AI-generated images. Instead, I suggest that we ought to consider mass AI-art as *bad art* rather than non-art.

3.2. Mass AI-Art

The term “AI-art”, in “mass AI-art” borrows from Boden and Edmonds (2009) work on the taxonomy of generative art. According to Boden and Edmonds, Computer Generated Art (CG-

Art) is art that is made by a computer with minimal to zero human interference. Although there is some overlap, Nannicelli's use of the term mass AI-art is narrower than my use of the term promptography. Whereas I include all prompt-based diffusion model image generators, Nannicelli excludes models that are custom made by artists or that are unavailable to the general population. For lack of better phrasing, Nannicelli is concerned only with models that can be effectively used by pushing a button.

The term "mass" in "mass AI-art" borrows from Carroll's (1998) work on mass art. For Carroll, mass art tracks art that is accessible to the widest possible audience with the fewest possible barriers to entry. While Carroll has in mind the untrained consumer of art, Nannicelli applies this accessibility condition to the producers of mass AI-art. In other words, mass AI-art is done by unskilled individuals with no training in how image generators work. Taken together "mass AI-art" denotes images generated with little human interference by individuals untrained and unskilled in AI image generating technology.

3.2.1. The Nature Argument

The first of Nannicelli's two arguments, that which I am calling the nature argument, proceeds as follows. We have an aesthetic experience of mass AI-art. However, said experience is limited to the formal aesthetic qualities of the images in question. Such images are devoid of relevant non-aesthetic features that some scholars (Goldman 1995; Brady 2003) argue play a role in our aesthetic experience of artworks. For example, Nannicelli argues that it is unclear at best whether AI-generated images can be understood as having a history in the same way as uncontroversial artforms, such as painting.

However, this is not unique to mass AI-art. Nannicelli suggests that our aesthetic experiences of nature are also predicated on formal aesthetic qualities. Likewise, according to Nannicelli, it would be difficult to suggest what kinds of non-aesthetic properties—such as historical properties—might bear upon our aesthetic engagements with nature. Nannicelli suggests that we might adopt a kind of strong formalism to understand our aesthetic engagements with mass AI-art. Nannicelli points to the extreme formalism of Zangwill (2013) that holds that aesthetic enjoyment and judgement of inorganic nature relies solely on perceivable physical appearances as a possible source of inspiration.

3.2.2. The Intention Argument

The second of Nannicelli's arguments, that which I am calling the intention argument, takes a different approach. Nannicelli directly responds to Wojtkiewicz' (2023) claims that AI-generated images produced by diffusion models can meet two necessary conditions for being art. The first condition is that a work has to be produced by an agent with the intent of said work being an artwork. The second condition is that the work in question must be the result of that agent's intentional action.

Nannicelli argues that neither condition obtains by drawing an analogy with polaroid cameras. While Nannicelli does not deny photography as an artform, he points out that polaroid cameras were designed for consumers to capture a moment or a memory, not to make artworks. According to Nannicelli, users of polaroid cameras do not have an intention to produce an artwork, thus failing to meet the first condition. Nannicelli argues that the way AI image generators are developed and marketed are similar to polaroid cameras.

Furthermore, just as the polaroid camera user only has to press a button, all the diffusion model user has to do is input a short prompt. Nannicelli takes this to be insufficient to be a case of intentional action. Here we might think of Scruton's (1981) claim that photography is causal and thus not an artform. While Scruton might be incorrect about photography tout court, Nannicelli holds that it is correct to attribute this claim to the use of polaroid cameras—and by extension AI image generators.

3.3. Counter Arguments

In this section I respond to both the nature argument and the intention argument. I put forth the suggestion that rather than denying art status to mass AI-art, we might instead think of it as being an example of *bad art*.

3.3.1. Rejecting the Nature Argument

I present two objections to the nature argument. First, I argue that non-aesthetic properties do factor in our evaluation and appreciation of mass AI-art. Second, I argue that formalism is the wrong framework for understanding aesthetic experiences of nature, as non-aesthetic properties are relevant to those cases as well. For simplicity, I will stick to historical properties as my example of a relevant non-aesthetic property.

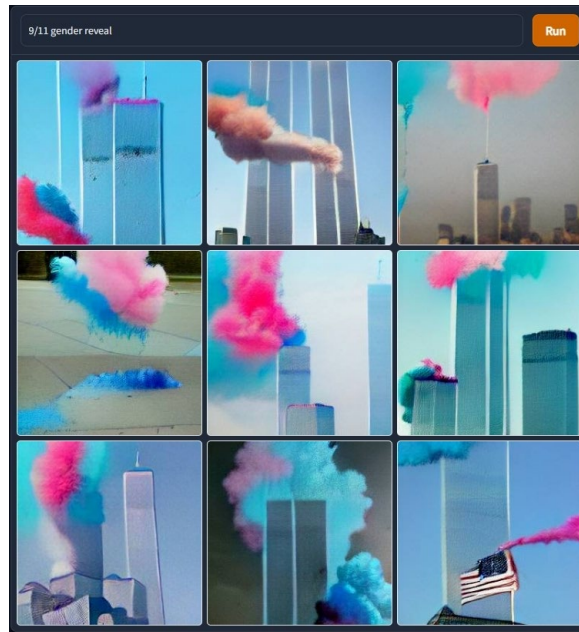


Figure 3.1: 9/11 gender reveal (Creator Unknown)

The prompts used in image generation seem to be an obvious example of a historical property that guides our experience of mass AI-art. Images posted online are often labeled with the prompt that led to the image’s creation. Such labels inform our understanding and evaluation of mass AI-art. Take *9/11 Gender Reveal* (Figure 3.1) for example. While it can be read as a shocking and provocative image, having the label *9/11 Gender Reveal* gives us the additional context of the joke/criticism being made.

The image can easily be interpreted as making light of a tragic situation—9/11—by juxtaposing it with the trivial practice of “gender reveal parties.” However, we might also grant a more subversive reading of the image that would understand it as a *critique* of gender reveal parties—which have become increasingly dangerous and extravagant—by comparing them to the 9/11 terrorist attacks. The consideration of a historical property of a work of mass AI-art here shows that our experiences of mass AI-art are transformed when we consider certain non-

aesthetic properties. Thus, contra Nannicelli, non-aesthetic properties *do* factor into our engagements with mass AI-art.

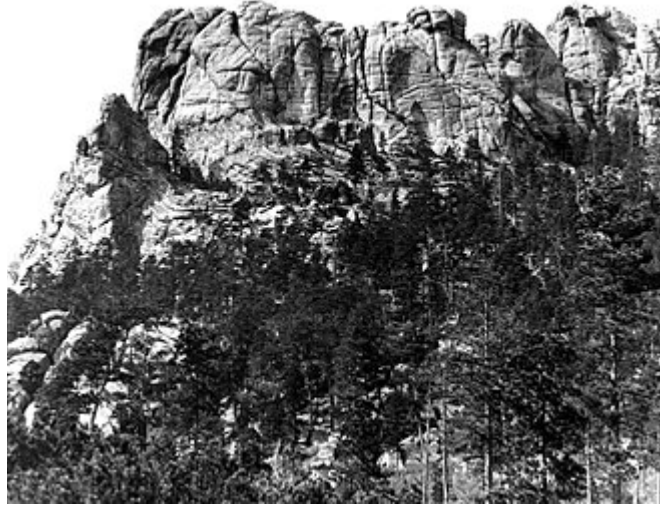


Figure 3.2: Six Grandfathers

Another way we might reject the nature argument is by denying that a strong formalist approach is the best approach for understanding our aesthetic experiences of nature. We might consider the example of “Six Grandfathers” (Figure 3.2) the mountain upon which “Mount Rushmore” was built. While it is possible to appreciate only the formal aesthetic qualities of the mountain, the site also holds spiritual and cultural significance for the Lakota Nation. Stories about the history and centrality of “Six Grandfathers” to Lakota culture serve as historical properties that no doubt enhance one’s aesthetic experience of the mountain. Given that these non-aesthetic properties inform how we view the mountain—as either a pretty rockface or a spiritual site—it cannot be the case that our aesthetic engagements with nature merely involve formal aesthetic properties.

Even if it were the case that non-aesthetic properties had no bearing on our aesthetic experience of mass AI-art, the comparison to our aesthetic experiences of nature seems misplaced. Nannicelli argues that when we engage aesthetically with mass AI-art, we only engage with its formal aesthetic properties. He further argues that this is how we engage with nature. Ergo, mass AI-art is more similar to nature and thus is not art. However, I have shown that we also engage with non-aesthetic properties of mass AI-art (such as the prompts used to generate such works). Furthermore, when we engage with nature, we also consider non-aesthetic properties in our evaluations in addition to formal aesthetic ones. With all of this in mind, we can confidently reject the nature argument.

3.3.2. Rejecting the Intention Argument

I present three objections to the intention argument. First, there are cases where polaroid photos are presented as art. This complicates the analogy that Nannicelli draws between polaroids and AI image generators. Second, even if we grant that Nannicelli is right about polaroid cameras and their users, we can see a disanalogy between polaroid cameras and diffusion models. This is because AI image generators are clearly understood by both the developers and the users as machines that assist in artmaking. Finally, there is more to both polaroid photos and AI image generation than merely pressing a button.

In the 1980s, Andy Warhol took a series of polaroid photographs to be displayed as artworks in galleries. Many of these photographs are highly valued, and still in circulation today. This is but one example among many of polaroid photos that have been presented and received in formal art settings such as galleries and museums. It is not the case that polaroid cameras are not used as an artmaking tool. For Nannicelli to say that AI image generators are like polaroid cameras insofar as neither machine is involved in artmaking practices is disingenuous at best.

However, Nannicelli could respond by saying that such examples of polaroid art are the exceptions rather than the rule. This would be concordant with his argument as he does grant that *some* instances of AI-generated images could be artworks—but such images would not fall under the category of mass AI-art. Even if we grant that Nannicelli is correct about polaroid cameras not being marketed as artmaking tools and most polaroid photos not functioning as artworks, his analogy between polaroid cameras and AI image generators does not hold. Many casual users of AI image generators will refer to themselves as “AI artists,” and a major selling point of these technologies is the democratization of art.

In this way it seems as though Wojtkiewicz’ first condition—that the work be produced by an agent with the intent that it be an artwork—is met. Rather than merely enjoying the generated image qua image, many users will flag their images as instances of AI art. While polaroids might be designed for capturing a memory on film, AI image generators are marketed and designed as machines that allow all users to participate in artmaking.

Furthermore, while we might think of the polaroid camera and the AI image generator as merely causal, we would be wrong to make this assertion. The lighting, the staging, and the angle at which the photo is taken are all important concerns for the polaroid photographer. The person behind the camera might end up taking several photos of their subject(s) to get the shot just right. Likewise, it is often the case that users of AI image generators will modify their prompts several times before selecting the output that they feel best captures their intentions.

Even in the case of a user inputting a single prompt and selecting the first image generated, there is still an intentional decision being made with respect to the choice of prompt and which image(s) to keep and which to discard. The resulting outputs are thus the result of the

agent's intentional actions, minimal though they may be. Thus, Wojtkiewicz's second condition—that the work in question be the result of the agent's intentional action—is met as well.

Having provided several reasons for rejecting the intention argument, I now turn to a discussion of bad art to set the stage for an alternative perspective on how to understand the phenomenon of mass AI-art.

3.4. Bad Art

Definitions of art can typically be classed into two types: There are *evaluative definitions* and *descriptive definitions* (G Dickie 2000). The former carry with them an evaluative component—*aesthetic value* might be a necessary condition for art status, for example. In evaluative accounts there is generally no distinction made between good art and bad art. Rather, there is art and non-art. Perhaps the most famous example of an evaluate account is Kant's ([1790] 2009) notion of fine art needing to be both original and exemplary of the form.

By contrast, descriptive definitions define art without an evaluative component. An example would be something like Danto's (1964) view that art becomes such when it is presented to an *artworld*—a community consisting of artists, art critics, art historians, art enthusiasts, etc. Descriptive definitions leave open the possibility of there being a distinction between good art and bad art.

I explore two ways of understanding the phenomenon of bad art. I first consider the *failed intentions view*. According to this view, bad art is the product of the failed intentions of the artist. I then consider what I call the *failed organic whole view*. According to this view, some properties of aesthetic works have a negative contribution to their overall value, leading to works of bad art.

While the failed intentions view has explanatory power in many cases, I show that it is insufficient for explaining mass AI-art as bad art. However, I will argue that the failed organic whole view is capable of giving an account of mass AI-art as bad art. While there are doubtless other ways of conceiving of bad art, I focus on these two views.

3.4.1. The Failed Intentions View

Intentionalism is the view that the meaning of an artwork is—at least in part—constituted by the intentions of its artist (Carroll 1992). If I intend for my work of painting to be understood as a work of abstract expressionism, and I succeed in my intentions, then the painting *succeeds* at being understood as a work of abstract expressionism in part because I succeeded in my intentions. The failed intentions view of bad art holds that bad artworks are the results of failed intentions. If it had been the case that I intended for my work of painting to be in the style of abstract expressionism, and I fail in my intentions—the painting looks like a landscape painting instead and people take it up as such, for example—then the painting *fails* as a work of abstract expressionism because I failed in my intentions. The failed painting is thus a bad work of art qua abstract expressionism.

This view doesn't quite work. In the case of mass AI-art, it is not the case that the prompts are always failing in their intentions—minimal though they may be. When a user gives the prompt “Batman fighting Godzilla” and receives an image depicting Batman fighting Godzilla that is to their satisfaction, it strikes me as though we ought to say that they have succeeded in their intentions. While there might be cases of failed intentions where the output of the prompt does not match the intentions of the user, the existence of success cases forecloses the possibility that all instances of mass AI-art are instances of failed intentions. As such, the failed intentions view cannot give an explanation of mass AI-art as bad art.

However, we have reasons to reject the failed intentions view. As has been pointed out by Strohl (2022), an artist can intend to produce a work of art that is bad. For example, say that I wanted to make a movie in the style of a bad horror film. I get together with some friends and we shoot a slasher film that is intentionally incompetent as a film. Despite succeeding in my intentions—to create a bad work of art—I have still created a bad work of art.

One could argue that while the movie I made with my friends is bad qua film, it might have other kinds of value. For example, it might be so incompetent that it succeeds as a work of comedy. Thus, while the film is bad qua horror, it is good qua comedy. This need not be a problem, however, as we can imagine cases of art where the art in question is bad simpliciter.

For example, I might choose to make a drawing of a house in the style of stick figures akin to what a three-year-old might draw. While we might praise the work of such a child for its demonstration of creativity and of what they have learned about the world, we would not praise the work on the basis of it being a good work of art. We might not even consider the child's work to be an example of art proper if they lacked the intention for the drawing to be understood as art.

If I were to draw such an image, we likely wouldn't praise my creativity, as I ought to know more about the world and the art of drawing than a three-year-old. However, so long as I was intending for my work to be received as art and I intentionally brought the work about, we could plausibly consider it to be a work of art. Nevertheless, we would identify my drawing as a *bad* work of art. Given that I can intend to produce a bad work of art and succeed at doing so means that the failed intentions view cannot account for all cases of bad art.

3.4.2. The Failed Organic Whole View

One way in which we can understand art is as an *organic whole* (G. Moore 1951). In this view, the various properties of a work of art all contribute to the value of the work as a whole.

Properties can be both aesthetic and non-aesthetic (K. L. Walton 1970). If all of the properties come together to create a work of value, then it is a good work of art. If some of the properties contribute negatively to the work as a whole, then we can say that the work is comparatively worse than a work whose parts all contribute positively to the whole.

For example, we might imagine a film with a great script, impressive special effects, good directing, and quality cinematography marred by poor acting performances. We might still consider the work to be a good film, but we would not consider it to be a perfect film.

Furthermore, we might comment on the poor acting performances as a negative aspect of the film that might impact our desire to recommend the film or engage in repeat viewing of the film.

If there are several properties of a work that contribute negatively to the whole, then we can say that the work is a bad work of art. Consider another film except this time, in addition to poor acting performances, the script is poorly written, the special effects are bad, the directing is incompetent, and the cinematography is a mess. We might say that this is a bad film, and a bad work of art.

It might also be the case that non-aesthetic properties contribute to our aesthetic experience of a film. Consider a movie with the disclaimer that “no animals were harmed during the making of this film.” However, it is later revealed to us that many animals *were* harmed in the making of the film. This historical property or fact of the film’s genesis would likely lead to a more negative evaluation of the movie than if no animals had been harmed in its making.

While there are cases of films that are “so bad they’re good,” (Dyck and Johnson 2017; Algander 2024), this does not seem to be what is going on in the case of mass AI-art. Many people seem to genuinely enjoy the aesthetic qualities of AI-generated images. Even Nannicelli concedes that works of mass AI-art produce aesthetic experiences and might possess some kind of aesthetic value.

3.5. Mass AI-Art as Bad Art

It is straightforward to see how works of mass AI-art constitute examples of a failed organic whole. Works of mass AI-art seem to have aesthetic properties that contribute positively to the value of the work as a whole. However, when we consider the non-aesthetic properties of works of mass AI-art we can see that they negatively contribute to the value of the work as a whole. For example, we might see the unskilled prompting—perhaps only a single short prompt without modification was used to generate the image—as a lazy artistic practice. Furthermore, we might see the fact that many users are completely untrained in how these models work as a sign that these users are ignorant of their artform. These are both historical properties of works of mass AI-art that contribute negatively to the whole. When we take these non-aesthetic properties into account, we have a way of understanding works of mass AI-art as failed organic wholes.

As it turns out, we have additional empirical evidence that can support this view. Many studies have been performed where participants are shown a variety of images—some of which are made by humans, some of which are made by AI. When told which images were AI-generated, people on average tending to give those images lower ratings (Hong and Curran 2019). While some mass AI-art is indistinguishable from human made art, when the non-aesthetic properties of such works are taken into consideration, many people reconsider their

assessments of the works and reevaluate them as having lesser value overall. One way this phenomenon can be explained is by understanding works of mass AI-art to be bad artworks.

Alternatively, it could be the case that people—in particular the ones who have participated in studies evaluating AI-generated images qua artworks—are merely *biased* against AI art as a legitimate form of art. I explore this idea elsewhere (Vlaad and Paul Forthcoming) and suggest that we ought to adopt the aesthetic virtue of *open-mindedness* when engaging with works of AI art. While this *bias interpretation* of the study results is at least plausible, it is not my aim here to defend the *bad art interpretation* against it. Rather, I intend to show that *all else being equal*, it is more plausible that mass AI-art is bad art rather than non-art.

If works of mass AI-art were non-art similar to nature, we would likely see a different outcome. When people learn that a given artefact they thought was a human made artwork is actually a work of nature, we tend not to value the overall aesthetic experience less. However, we certainly reframe our understanding of the artefact in question. For example, we might think that rocks displayed in an intricate pattern were once the work of a creative artist. Upon learning of their genesis we might then deny that the pattern of rocks is creative (Paul and Stokes 2018), but it seems unlikely that we would allow this revelation of the pattern's non human origins to lessen our overall aesthetic evaluation of the rocks tout court.

Both in the case of the intricate pattern of rocks and the case of mass AI-art it is non-aesthetic features that decide our overall aesthetic evaluation of each case. On the one hand, the rocks lack a historical property of being made by a human artist. On the other hand, mass AI-art possesses a historical property of being generated through the unskilled and untrained use of AI. Whereas in the former case the non-aesthetic property does not negatively contribute to our

aesthetic evaluation of the pattern of rocks, in the latter case it does seem to negatively contribute to our aesthetic evaluation of mass AI-art.

Furthermore, there are some who hold that there are aesthetic experiences of value that can only be found in nature. Kant's ([1790] 2009) articulation of the sublime is perhaps the most famous example of such an experience. All of this is to say that we tend not to consider aesthetic experiences of nature to be of lesser value than aesthetic experiences of art. We simply consider the two to be different aesthetic experiences.

In this chapter I have given reasons for rejecting Nannicelli's account of mass AI-art as non-art. Furthermore, I have provided my own account of mass AI-art as bad art by understanding works of art as organic wholes. I have provided reasons for why my view better explains our reactions to mass AI-art than Nannicelli's. When taken together, we can see that mass AI-art is better understood as a kind of bad art, rather than a kind of non-art. In the next chapter I explore whether AI constitutes a new artistic medium and by extension a new artform.

Chapter 4: Misalignment

4.1. The Aesthetics of AI Art

Some scholars have begun to argue that we ought to understand AI as a new artistic medium with unique properties (Young and Terrone 2025; Helliwell 2025). Others have discussed how we might identify general styles of AI art and individual styles of AI artists (Cross 2024). Very little has been written about what these styles might look like in terms of what kinds of unique aesthetic value(s) they might possess. There are a few exceptions, with scholars such as Helliwell (2023) identifying the aesthetic feature of *weirdness* and Rini (2024) introducing the concept of *the artificial sublime*. In this chapter I take up questions of style and aesthetic value as they pertain to AI art.

Helliwell's (2023) account identifies two aesthetic qualities common to visual AI artworks—*weirdness* and *convincingness*. The former relies on the *failure* of AI systems to replicate features of works made by humans—such as the famous example of models producing images of hands with too many digits. The latter relies on the *success* of AI systems to replicate features of works made by humans—such as Boris Eldagsen's *The Electrician*, which many took to be a photograph rather than an AI-generated image. I take both of these qualities to belong to distinct *styles* of visual AI art.

While weirdness was the common style when Helliwell was writing her dissertation—which primarily focused on the use of GANs to make art—contemporary AI image generation leans more towards convincingness. I argue that this is not merely a case of “technological progress,” but a design choice on the part of AI developers. By understanding weirdness as a

kind of failure in need of correction, developers have implicit aesthetic judgements. Namely, that weirdness is an undesirable aesthetic quality and should be mitigated.

AI developers discuss fixing such “mistakes” as maximizing the alignment between the semantic meaning of textual prompts and the images generated by said prompts (Hessel et al. 2022; J. Li et al. 2022, 2023; Singh and Zheng 2023). While these “errors” are mostly discussed in terms of aesthetics, implicit—and sometimes explicit—ethical judgements are also being endorsed. This manifests in at least three ways.

First, some developers insist that AI art must look *natural* rather than *unnatural* (Theis 2024; Narasimhaswamy et al. 2024). The images generated must not contain hands with too many fingers or distorted faces. This value is also referred to as *perceptual quality* and denotes “...the extent to which [a constructed image] \hat{x} looks like a valid natural image” (Blau and Michaeli 2018, 1). The idea here is that AI-generated images that are indiscernible from those produced by means other than AI are of a higher quality.

Second, some developers insist that AI art must be *beautiful* rather than *ugly* (Chen et al. 2024). The images must not be gross or discomfoting to look at. AI-generated images computer scientists deem to be of high quality possess what they term “a strong sense of aesthetics” (Wang et al. 2024, 1). These same researchers argue that novice prompt givers often have trouble knowing which text inputs will produce “successful” outputs. This is framed as part of the alignment problem. To overcome this issue the authors present *PROMPTCHARM*, a system that *automatically* corrects user prompts such that the images produced will have greater aesthetic value as determined by the developers.

Finally, some developers insist that AI art must be *inoffensive* rather than *offensive* (R. Li et al. 2023; Petsiuk and Saenko 2024). The images generated should not create controversy or deal with polarizing subject matter. Developers are working to eliminate outputs that express bias and harmful content—though they do not differentiate between content that is *harmful*, and that which is *offensive*. Offensive content might take the form of social commentary that makes some people uncomfortable, whereas harmful content might be that which perpetuates racist and sexist bias or is sexually exploitative. While harmful content is likely to always be offensive, the converse does not hold. AI developers do not make this distinction and instead consider offensive content to be either a subset of or coextensive with harmful content.

According to Helliwell (2023), if we limit the discordance present in prompts and AI-generated images, we miss out on at least one unique aesthetic value of AI artworks (weirdness). While it is true that by limiting this discordance the average user will be more likely to generate images that rank highly on alignment as articulated by developers, skilled prompters will still be able to exploit AI systems to produce works with lower alignment.

I suggest that AI artworks with lower alignment possess a unique aesthetic feature that I call *misalignment*. This occurs when there is a big enough discordance between the most plausible interpretation a reasonable person would have of how a prompt ought to manifest and the output that is actually generated by AI. As the most plausible interpretation a reasonable person would have of human hands is that they possess five digits, an AI-generated image of hands with extra digits clashes with our expectations and counts as an example of the aesthetic feature of misalignment. The weirdness identified by Helliwell is a subtype of misalignment in this proposed account.

This chapter proceeds as follows. I begin with a brief overview of Helliwell’s notion of weirdness as a form of *norm violation* by way of *non-human failure*. I reframe this notion of weirdness in terms of misalignment. Following Cross (2024), who suggests that we might identify unique styles of AI art by looking at different prompting styles used by artists, I identify a style of AI art that I call *bias-prompting*. I suggest that bias-prompting involves the manipulation of prompts to generate images that clearly illustrate the presence of *algorithmic bias* (Danks and London 2017; Fazelpour and Danks 2021; Stinson 2022) in diffusion models. I present bias-prompting as a more intentional strategy for developing images that possess the aesthetic feature of misalignment. I conclude with a defence of misalignment against technological “improvements” suggested by AI developers.

4.2. Weirdness

In this section I give two ways for understanding the aesthetic feature of weirdness common to some AI art. I first give Helliwell’s account. I then critique Helliwell’s account and provide an alternative way of understanding weirdness.

4.2.1. Weirdness as Non-Human Failure

According to Helliwell (2023), there is a weirdness common to many works of visual AI art. Drawing on Carroll’s (1999) work on horror and humour, Helliwell describes such weirdness as an act of *norm violation*. The images in question serve as violations of established norms, categories, and expectations. For example, early AI art that featured human hands with too many digits or blurry faces instead of eyes, a mouth, and a nose, violate our expectations of what humans are supposed to look like.

Helliwell goes one step further and suggests that the norm violations present in AI art are distinct from those found in other art forms. According to Helliwell, these particular norm violations are constituted by non-human failure. The argument goes like this: a human being drawing a cat might make the mistake of getting the proportions wrong, but they would not accidentally draw a cat without any limbs. This is because humans possess the relevant schema of a cat. We know what cats are supposed to look like, even if we are not all experts at drawing them accurately.

The mistakes made by AI are non-human mistakes. AI does not possess the relevant cat schema, and as such is just as likely to depict a cat without any limbs than it is to get the proportions wrong. Helliwell's point here is that AI uses different conceptual building blocks than humans do. Whereas (most) humans possess the concept of a cat, AI does not.

Helliwell draws on an example from computer-vision research wherein an image of a cat was overlaid with elephant skin texture. Whereas humans were able to correctly identify the image as a cat, the AI incorrectly identified the image as an elephant. This was the case because unlike humans, the AI prioritized texture over other salient features such as shape. Thus, when AI fails at identifying or generating an image of a cat, it does so in ways that are distinctly non-human. The particular weirdness of AI art is derived from the specific ways in which AI engages in non-human failure.

We might consider the AI-generated nude portraits of Robbie Barret. While one can make out the fact that these are intended to be nudes of the human body, the proportions are such that it is difficult to tell if there are one or multiple subjects in the images. Furthermore, the subject(s) are depicted as disturbing abstract blobs of flesh with no clear demarcation between the body and the environment. These are norm violations that are the result of AI failing to correctly interpret a

prompt. The ways in which the AI fails to correctly depict the human body are examples of non-human failure.

4.2.2. Weirdness as Misalignment

An alternative way of understanding the weirdness of some AI-generated images is by thinking of them as instances of misalignment between human expectations of how prompts should be interpreted and how AI actually interprets prompts. The key difference with this approach is that the results are not conceptualized in terms of failure. While such images do seem to constitute acts of norm violation, such acts should not be understood as being failures.

In articulating the distinction between cats drawn badly by humans and the weird images generated by AI, Helliwell suggest that the difference is “between intending to draw a cat and failing, versus arguably having no intention to depict a ‘cat’, merely the goal of reproducing cat images” (2023, 252). If AI cannot be said to be intending to generate an image of a cat, then it cannot be said to be failing at this goal. It could be the case that the AI is performing very well by some other metric, and if we want it to make more realistic images of cats then we ought to design the model differently.

When understood not as failure, but as a difference in expectations in how a prompt ought to be interpreted, we can articulate the norm violations present in certain AI artworks as being caused by misalignment. The most plausible interpretation a reasonable human would have of the prompt “orange cat” would be an image of an orange creature with pointy ears, a tail, and four legs. AI does not always interpret the prompt “orange cat” in a way that is concordant with human expectations. When this occurs, AI developers would say that there is a low level of alignment between the prompt and the output.

Misalignment is an aesthetic feature that is unique to AI. This is in part due to the stochastic nature of such machines. By acquiescing some degree of autonomy to the model, human users open the door for misalignment. Young and Terrone (2025) describe this feature of AI as *recalcitrance*. That is to say that AI presents users with a form of resistance that must be negotiated. Whereas standard painting has *inert* recalcitrance insofar as the painter is always in control of her tools and materials, AI has *dynamic* recalcitrance insofar as there will always be some degree of control that is lost when using AI. While there are painting techniques that relinquish control, such as action painting, the dynamic recalcitrance in this case is not essential to painting in the way it is to AI. Paintings cannot possess misalignment as a result of recalcitrance, whereas it is a common feature of AI art.

Helliwell expresses a concern that with new technological developments that strive to enable AI to more consistently produce more “successful” images we will lose out on the unique weirdness characteristic of AI art. While we might have fewer instances of weirdness resulting unintentionally from AI, in the next section I argue that skilled prompters are finding new ways to prompt AI to generate images that possess the aesthetic feature of misalignment.

4.3. Bias-Prompting

In this section I follow a suggestion made by Cross (2024) that we might be able to discern both the individual style of particular AI artists and general styles of AI art by looking to the prompting strategies being deployed. I identify a new general style of AI art that I call bias-prompting. I claim that this style involves the manipulation of prompts to expose algorithmic bias present in AI to produce the aesthetic feature of misalignment. When engaged properly, these images serve to critique systemic injustices such as sexism and racism by revealing the biases present both within generative models as well as society in general. First, I explain the

concept of algorithmic bias. I then give my account of how bias-prompting produces images that possess the aesthetic feature of misalignment by considering a concrete example of bias-prompting in the form of *In/Visible*, an AI art exhibit curated by Linda Dounia.

4.3.1. Algorithmic Bias

Algorithmic bias is an umbrella term used to describe the various ways that computational systems—especially autonomous ones—can express bias in their processing and outputs (Danks and London 2017; Fazelpour and Danks 2021; Stinson 2022). While the term is sometimes used in a value neutral way, many instances of algorithmic bias end up reifying dominant oppressive structures (Noble 2018). For example, some systems used to predict recidivism rates of convicted offenders have been found to have racial biases (Angwin et al. 2022). Of particular relevance in the case of AI-generated images is the phenomenon of training set bias. When a certain class of items are overrepresented in an algorithm’s training data, we say that there is a sampling bias in favour of that class of item. This might mean that English language tokens are overrepresented in the training data for an LLM, leading to a bias in favour of English and against languages other than English.

AI image generators are trained on a vast quantity of text-image pairs. Despite the amount of diverse training data, research has shown that many models exhibit racial and gender biases in their outputs (Agarwal et al. 2021; Bianchi et al. 2023). Furthermore, the art historical datasets used to train contemporary models tend to have a bias towards Western European art (Mazzone and Elgammal 2019). Naturally, this limits the kinds of artworks that can be produced by prompting diffusion models. This confluence of factors often leads to the exclusion and erasure of women and people of colour from the outputs generated from these image generating

models. Furthermore, these models also reproduce harmful gender and racial stereotypes in their outputs leading to material consequences for marginalized communities.

I define bias-prompting as a style of AI art wherein models are prompted to generate images that showcase these instances of erasure, exclusion, and stereotyping. This is done in order to critique current systems of domination and configurations of power. Bias-prompting relies on a low level of alignment between the most plausible interpretation of prompts by a reasonable person and the actual images produced by AI. In this instance, the low level of alignment is achieved by exploiting algorithmic bias.

4.3.2. In/Visible

So far, the discussion of bias-prompting has been rather abstract. I now turn to a concrete example of bias-prompting in the form of the art exhibit *In/Visible*. Organized by Senegalese curator Linda Dounia, *In/Visible* is “...a way for Black artists to feel less lonely in their experience of A.I., to have their challenges expressed in a way that resonates materially and emotionally, to reject the normalization of their exclusion in emerging technologies” (Lawson-Tancred 2023). The exhibit brings Black artists using AI together to showcase artworks that underscore how algorithmic bias impacts how diffusion models (mis)represent Blackness.



Figure 4.1: Blonde Braids Study II (by Minnie Atairu)

Consider *Blonde Braids Study II* by Minnie Atairu (Figure 11). Despite the title suggesting that the figures in the image might possess blonde braided hair, the image instead depicts two figures with what appear to be golden ribbons in place of hair. Minnie Atairu, the artist responsible for generating the image, describes the work and its genesis as follows:

In ‘Blonde Braids Study II,’ I examine the ways in which a text-to-image algorithm — Midjourney (v4) — renders a portrait of Black identical twins adorned with blonde braids. The resulting image underscores significant gaps in the training data, which inevitably precipitates a flattened representation of the Black identity outlined in the text prompt. Instead, the algorithm generates an interpretation that evokes the semblance of fraternal twins, their hair styled in blonde, permed waves. (Feral File, n.d.)

The image alongside the curated description illustrates the “flattening” of Black representation that occurs due to the presence of algorithmic bias against dynamic Black subjects. When we consider the artwork, we are invited to reflect on the biases present within AI image generators, the datasets upon which they are trained, and the broader social biases, attitudes, and structures that lead to the perpetuation of harmful racial stereotypes in the first place.

These images possess the aesthetic feature of misalignment insofar as the model is unable to generate an image of Black subjects with blonde braids. While the most plausible interpretation by a reasonable person of a prompt detailing Black subjects with blonde braids would be an image of Black subjects with blonde braids, the output lies in contrast to this interpretation. Thus, a low level of alignment has been achieved in this case. The skilled prompter—in this case Atairu—is able to use the model to both generate an image with the unique aesthetic value of misalignment while concurrently critiquing both algorithmic bias and systemic racism.

4.4. The Case for Misalignment

I have thus far argued that misalignment is an aesthetic feature unique to AI art. I have further identified bias-prompting as a general style of AI art that requires some degree of skill. While some readers will agree with Helliwell and myself that the unique aesthetic features of AI art are worth preserving—be they weirdness in particular or misalignment in general—others might remain unconvinced. After all, it seems plausible to suggest that we would want to limit the presence of algorithmic bias present in AI systems regardless of whether doing so would reduce the possibility of generating images possessing misalignment.

However, it is unclear whether maximizing alignment would achieve this goal. Additionally, the ways in which alignment has been defined by AI developers is *itself biased* towards images with particular aesthetic values. For example, scholars such as Brady (2011) have discussed the possibility of *negative aesthetic value*. Just as we might have a meaningful aesthetic experience of a beautiful sunset, so too might we have a meaningful aesthetic experience of an ugly rotting tree. Smuts (2009) has discussed the value of negative affect in art with reference to both scary movies and tragic characters, whereas Strohl (2012) has discussed negative aesthetic experience in the horror genre in detail. All of the AI developers discussed completely overlook the value of negative aesthetics and assume that only positive aesthetic value is worthwhile.

We ought to be concerned about developers adopting the singular goal of maximizing alignment for two reasons. First, limiting the types of aesthetic values—unique or otherwise—that can be generated through the use of AI functions as a kind of censorship. For example, by conflating the concepts of harmful and offensive and attempting to limit images considered as such would prevent artists from using AI to create works of parody or satire that criticize current configurations of power and domination. Mel Brooks famously satirized the nazis in his film *The Producers* through the musical number “Springtime for Hitler.” Such social commentary would not be possible with AI if maximizing alignment includes limiting the generation of content deemed offensive by the developers.

Second, a singular focus on maximizing alignment not only *limits* the kinds of art one can make with AI, but it also *shapes* the kinds of art that will be made. If watercolour paints were widely unavailable and acrylic paints were sold at very low prices in every store, we would likely see a large increase in works of acrylic painting and a severe decrease in works of

watercolour painting. This in turn would shape how both acrylic paintings and watercolour paintings are understood and evaluated by artists, critics, and the general public. To use an example from cinema, if we suddenly stopped making horror films, this would be seen as a loss. But if we only made horror films, this too would be seen as a negative.

By focusing on maximizing alignment in AI systems, developers risk imposing a particular set of artistic values at the cost of both unique aesthetic values—such as weirdness and the broader misalignment—and aesthetic diversity in general. We should remain skeptical of the framing of these technological developments as “progress” rather than the imposition of a different set of biases.

In the next chapter I explore whether AI constitutes a unique artistic medium, and by extension an artform.

Chapter 5: AI, Medium, and Artform

5.1. Tool, Collaborator, or Something Else?

Whether generative AI should be understood as a tool used to make art or as a kind of artistic collaborator remains an open question. Considering AI as merely a tool doesn't seem to capture the degree of autonomy exerted by the machine. However, considering AI as an artistic collaborator seems to grant too much agency to the machine. Some scholars have argued in favour of the tool view (Wojtkiewicz 2023), whereas others support the collaborator view (Khosrowi et al. 2025). Young and Terrone (2025) attempt to circumnavigate this dichotomy by proposing an alternative option—that AI is an artistic medium. Helliwell (2025) endorses a version of this position, and further argues that in addition to constituting an artistic medium, AI also constitutes an artform. Cross (2024) provides a different view and conceives of AI art as a kind of participatory art wherein AI functions like a participant in a work of performance art.

In this chapter I defend a modified version of Helliwell's account. I begin with a discussion of the concept of artistic medium. I then provide a reconstruction of Young and Terrone's argument for AI as artistic medium. I show that it is a category-mistake to consider AI as the artistic medium. Instead, I argue that we ought to consider AI the vehicle through which the artistic medium of AI art is expressed. I then give an overview of Helliwell's arguments for AI as both an artistic medium and an artform. I suggest that we should take my use of AI art to be equivalent to a charitable interpretation of Helliwell's use of the term AI. Finally, I consider a counter argument in the form of Cross's view of AI art as a kind of participatory art and subsequently defend a modified version of Helliwell's view of AI art as a unique artform.

5.2. Artistic Medium

Scholars who discuss artistic medium typically fall into two camps. The first camp holds that the concept of artistic medium is unhelpful at best, and outright incoherent at worst (Beardsley 1958; Carroll 2003). The second camp holds that at the very least the concept of artistic medium helps to clarify issues in the ontology of art—especially with regards to the classification and evaluation of artworks (D. Davies 2009; Lopes 2014). Helliwell (2025) falls firmly in the second camp. I shall assume for the sake of argument that the concept of artistic medium has useful explanatory power.

Davies (2009) draws an important distinction between the *physical medium*—or the *vehicle*—of a work, and the artistic medium.

According to Davies, the vehicle consists of the tools and materials used to construct a work of art and the artistic medium consists of the intentional ways in which an artist uses the tools and materials to make artworks. In other words, the vehicle is the means by which the artistic medium is transported or expressed. For example, we might say that the paintbrush, the canvas, and the acrylic paints are the physical medium of a work of painting, whereas the artistic medium of a work of painting is the set of practices applied to the materials and tools to express a given work of painting.

Consider the example of painting one's living room walls. Imagine two scenarios. In the first scenario, a person decides to paint their walls white so that the room appears brighter and more inviting. In the second scenario, the person decides to paint a mural depicting some sort of landscape. In both scenarios, the vehicle is the paint and the walls. However, only the work in the second scenario would be considered to be in the artistic medium of painting. This is because of the techniques and practices that have been applied to the walls and paint (certain brushstroke

techniques, depicting a landscape, etc.) The ways one applies the relevant techniques and practices to the tools and materials constitutes the artistic medium.

Gaut (2010) endorses the *medium-specificity thesis* in his analysis of the concept of artistic medium. The thesis roughly states that for a given work of art to belong to a given artistic medium, two conditions must obtain. First, some evaluations of the work must make reference to a particular artistic medium. Call this condition *medium-specific evaluation* (MSV). Second, some explanations of a work and its artistic properties must make reference to a particular artistic medium. Call this condition *medium-specific explanation* (MSX). Additionally, Gaut proposes a third condition that must obtain for a given artistic medium to also be considered an artform. This condition states that iff a particular medium has unique artistic qualities, then it constitutes a unique artform. Call this condition *medium-specific artform* (MSF). It is not the case that artforms cannot share any artistic qualities. Nor is it the case that all artworks of a given artform instantiate the artistic qualities unique to the artform, Rather, for a given artform to be unique it must possess the capacity to instantiate some artistic qualities that other artforms do not.

Consider the example of painting. When we evaluate a given work of painting we might compare it to other paintings in similar styles. Furthermore, we might make reference to the brushstroke techniques used by the painter. For example, we might evaluate a Pollock painting in terms of how well the technique of action painting was executed. These evaluations make reference to the artistic medium of painting, ergo MSV obtains in this case.

When we explain a given work of painting, we might situate it as responding to a history of other paintings or locate it within a particular school or tradition. Additionally, we might discuss why certain paints or techniques were applied to achieve a particular outcome. For example, we might explain works of surrealism as responding to earlier movements in painting

that attempted to represent the world as it presents itself to us. These explanations make reference to the artistic medium of painting, ergo MSX obtains.

Finally, we can see that painting has unique artistic characteristics insofar as it involves the application of paint to a surface using specific techniques and brushes that we do not see in other artforms. Ergo, MSF obtains and painting is an artform. With all of this in mind, we can understand painting as both an artistic medium and an artform.

5.3. AI and Artistic Medium

In this section I reconstruct Young and Terrone's argument that Midjourney constitutes a new artistic medium unlike other artistic media. This is because according to Young and Terrone, AI is most like gardening. After summarizing this view, I draw on Davies distinction between the artistic medium of an artwork and the physical medium or vehicle through which it is expressed to show that Young and Terrone have committed a category-mistake by considering Midjourney to be the artistic medium rather than the vehicle.

5.3.1. Midjourney as Artistic Medium

Young and Terrone draw an analogy between Midjourney and the medium of gardening. They begin by exploring the role of the gardener in the making of a garden. The gardener plants seeds, waters flowers, and monitors the garden to make sure its plants are getting enough sun and nutrients. Borrowing terminology from Spinoza, they use the phrase *natura naturans* to describe "what a garden is doing," and the phrase *natura naturata* to describe "what a garden has done." The gardener works alongside *natura naturans* (soil quality, weather patterns, the particular ways that different plants grow) to produce a work of *natura naturata* (a garden).

Young and Terrone draw an analogy between the uncertainty of *natura naturans* and the stochastic qualities deployed by Midjourney. If prompts are like seeds, then Midjourney's stochastic processes are akin to the uncertain variables of *natura naturans*. Young and Terrone propose the phrase *machina naturans* to capture this quality of Midjourney. So, in the case of Midjourney the prompter works alongside *machina naturans* to produce a work of AI art.

According to Young and Terrone, what differentiates Midjourney from other artistic media is the necessary presence of *dynamic recalcitrance*. This recalcitrance can be characterized as a kind of unpredictability and autonomy possessed by both *natura naturans* in the case of gardening and *machina naturans* in the case of Midjourney. Whereas many artistic media possess *inert recalcitrance*—that is to say that a tool such as a paintbrush is generally predictable in the hands of a trained painter—Midjourney possesses dynamic recalcitrance necessarily. While higher levels of recalcitrance are possible with painting (such as the action painting techniques of Jackson Pollack) in the case of painting recalcitrance is a possibility rather than a necessity. According to Young and Terrone, it is this necessary dynamic recalcitrance that partially constitutes Midjourney as an artistic medium.

While some scholars have argued that at least some gardens can be artworks (Miller 1993; Cooper 2003), it is not necessary that they be for Young and Terrone's analogy between Midjourney and gardening to work. What matters for Young and Terrone is that Midjourney is an artistic medium with similarities to gardening—whose status with respect to artistic media can be left indeterminate.

5.3.2. Midjourney as Vehicle

In *The Concept of Mind*, Ryle ([1949] 2002) gives an account of what it means to commit a *category-mistake*. This occurs when one treats a given object as belonging to one category when it in fact belongs to another category. Ryle is concerned with giving an account of how attempting to locate the mind is misguided and uses the famous analogy of the university to present his point.

Consider a visitor to the University of Oxford. The visitor is shown many locations: libraries, colleges, museums, playing fields, etc. The visitor then asks where the university is located. However, the university itself is not a particular college or library. Rather, the university is "...just the way in which all that [the visitor] has already seen is organized" (Ryle [1949] 2002, 16). It is a mistake to locate the University of Oxford in any of its constitutive parts. The university is instead comprised of all of these locations and organized under the banner of the University of Oxford.

Young and Terrone make the move of identifying Midjourney as the artistic medium in the case of AI art. This can be understood as a category-mistake. The practices (such as the style of prompting and the types of prompts used) applied to Midjourney are distinct from the model itself. We apply these techniques and practices to Midjourney. Recall that an artistic medium is defined as the ways one applies certain techniques and practices to a given vehicle. Consider the earlier case of painting a wall. It is the ways that particular techniques and practices are applied to the wall that make a mural a work in the artistic medium of painting. The tools and materials used to make the mural (the paint, paintbrush, and wall) are the vehicle through which the artistic medium of painting is expressed.

Identifying Midjourney as the artistic medium is equivalent to identifying the tools and materials used in the mural as the artistic medium, or, in Ryle's example, identifying a particular

college or library as the University of Oxford rather than as one of its constituent parts. Given that we apply techniques and practices (such as prompting) to Midjourney, it should be understood as the vehicle through which the relevant artistic medium is expressed. The artistic medium in this case is the techniques and practices applied to Midjourney (such as the style of prompting).

Consider the analogy with gardening. In the case of gardening, the seeds and natural elements are analogous to the vehicle, whereas the set of techniques and practices we call gardening is analogous to the artistic medium. In the case of AI, we ought to understand Midjourney as the vehicle and we ought to understand the set of techniques and practices applied to Midjourney as the artistic medium.

I shall refer to this artistic medium as *AI art*. I mean for this term to be as broad as possible to account for a variety of different forms and uses of AI to make art, such as what Anscomb (2025) calls *synthetic imaging* or what artists like Boris Eldagsen refer to as *promptography* (Williams 2023). While my focus here is largely on AI and visual arts, AI art also includes the use of LLMs to make AI-generated literary texts (which I discuss in the following chapter).

5.4. AI Art and Medium Specificity

Helliwell also makes the claim that AI is an artistic medium. However, Helliwell goes one step further and claims that AI constitutes an artform as well. By AI, Helliwell is referring to any machine learning system. While she is not referring to a particular model, such as Midjourney, she is still referring to the class of objects we call models. In this way, Helliwell is also committing a category-mistake like Young and Terrone.

However, Helliwell also makes reference to techniques and practices that we apply to models that are distinct from particular models (such as training one's own algorithm or skillfully prompting a model). Given this, we can rescue Helliwell's view by interpreting her use of the term AI to be more expansive than merely referring to machine learning systems. In other words, we can interpret Helliwell's use of the term AI to refer to the techniques and practices applied to models rather than to just the models themselves.

Note that this use of the term AI is equivalent to my use of the term AI art above. As the most common contemporary way of using the term AI is to refer to models (diffusion models, large language models, etc.) I will use the term AI art instead of AI to avoid confusing the artistic medium of AI art (which consists of techniques and practices applied to models) with the models themselves (which we refer to as AI). With this caveat in mind, I will argue that Helliwell's view can largely be rescued.

In this section I reconstruct Helliwell's argument for why we should take AI art to be an artistic medium and an artform. Helliwell applies the medium specificity thesis to the case of AI art, arguing that MSV, MSX, and MSF all obtain in this case. While I agree with Helliwell's assessment of AI art as an artistic medium, I challenge some of the artistic properties that she lists as being unique to the artistic medium of AI art.

5.4.1. AI Art and MSV

First, let us consider MSV. This condition obtains when our evaluations of artworks make reference to the particular artistic medium to which they belong. Helliwell suggests that when we evaluate works of AI art, we often make reference to the ways in which AI systems were used to make said works.

Helliwell makes the argument that when referring to works of AI art as artistic achievement, we often make reference to the proposed artistic medium of AI art. Helliwell uses the example of Anscomb's (2025) analysis of the work of Annika Nordenskiöld to make this point. In reference to the promptographs of Nordenskiöld, Anscomb discusses how Nordenskiöld is able to develop a clear style despite the stochastic nature of the AI systems with which she is working. We can also think of this in terms of the dynamic recalcitrance proposed by Young and Terrone (2025). Nordenskiöld works with and against the AI to produce images with a clear style that can be recognized as belonging to her.

Helliwell argues that Anscomb's reference to the stochastic nature (or dynamic recalcitrance) of AI in her analysis of Nordenskiöld's work constitutes an instance of making reference to the relevant artistic medium when providing an evaluation of a work produced in that medium (MSV). Helliwell takes this as evidence that MSV obtains in the case of AI art, and provides a list of other ways we might make reference to the artistic medium of AI art in our evaluations of AI art. For example, we might make reference to:

1. Deft prompting
2. Development and training of one's own algorithm
3. Recognising and/or developing characteristics that are particularly interesting or valuable
4. The effort to create coherence in something made by an AI (Helliwell 2025, 44)

For example, we might praise a particularly interesting AI-generated image produced from a clever or unusual prompt (deft prompting). We might also praise an artist for curating her own dataset and training her own model, such as was the case with Anna Ridler and her work *Mosaic Virus* (development and training of one's own algorithm). Furthermore, we might praise works

with bizarre aesthetic qualities, such as Robbie Barrat’s amorphous bloblike AI-generated nude portraits (Recognising and/or developing characteristics that are particularly interesting or valuable). Finally, we might praise an artist for developing her own style, as with the example of Nordenskiöld given above (The effort to create coherence in something made by an AI). In all of these cases praise is given by making reference to the set of practices and techniques applied to AI. In other words, these artistic achievements are evaluated by making reference to the artistic medium of AI art. Ergo, MSV obtains.

5.4.2. AI Art and MSX

Next, let us consider MSX. This condition obtains when our explanations of artworks and their artistic properties make reference to the particular artistic medium to which they belong. Helliwell suggests that when we explain works of AI art, we often make reference to the ways in which AI systems were used to make said works.

Helliwell makes the argument that explanations of works of AI art and their artistic properties often make reference to the limitations of AI systems. For example, AI is limited in terms of what it can represent by its training data. While this does not mean that AI cannot generate novel images, it does mean that it cannot vary wildly from its training data. For example, a model trained on images of impressionist paintings isn’t going to generate an image that resembles a work of cubism.

Helliwell further claims that AI can only make images that are *unreal*. By this Helliwell means that AI can only reference what is included in its training data and not “the real world.” We can contrast this with photography. Photographs necessarily have a direct connection to the real world insofar as the subjects of photography are artefacts and individuals that exist in the

real world. It is true that one can create the effect of a subject that does not exist using photography. For example, one could paint black spots on a white sheet and place it far in the distance to give the impression of a cow. However, even if the resulting photograph appears to be of a cow, there is a counterpart in the real world made of a painted sheet to which the photograph corresponds. There are no such real-world counterparts for AI-generated images, as AI is only referential to the dataset. It is in this sense that Helliwell uses the term unreal.

Helliwell goes on to make the point that explaining works of AI art, in particular their artistic properties is often done by making reference to the dataset. Consider the example of the model trained on images of impressionist paintings given above. We can explain why a given output of said model possesses artistic properties common to works of impressionist painting by making reference to the fact that it was trained exclusively on paintings with such artistic properties. Dataset curation falls under the umbrella of the set of techniques and practices applied to AI that comprise the proposed artistic medium of AI art. This means that explanations of AI art and its artistic properties at least sometimes must make reference to the particular artistic medium (AI art) to which it belongs. Ergo, MSX obtains.

5.4.3. AI Art and MSF

Finally, let us consider MSF. This condition obtains when artworks possess particular artistic properties that are unique to the artistic medium to which they belong. Helliwell suggests that works of AI art express artistic properties unique to the artistic medium of AI art.

Drawing on the work of Gaut, Helliwell makes a comparison to film. A work of film that possesses the particular aesthetic properties unique to the artistic medium of cinema can be said to be *cinematic*. Accordingly, works of film that possess several of these properties are more

cinematic than films that possess fewer properties. Some films might not possess any of these properties and are said to be *uncinematic*. Smith (2006) gives the example of the 1981 film *My Dinner with André* as a film that is uncinematic, as it is a recording of a conversation and little would be lost if it had been produced as a play rather than a film. While the presence of the unique artistic properties of cinema make a work of film cinematic, Helliwell suggests that we ought to understand the presence of the unique artistic properties of AI art make a work *machinic*.

Helliwell provides a list of potential candidates for artistic properties unique to AI art.

They are as follows:

1. Synthesize large amounts of data into a final output.
2. Produce artificial images (of unreal subjects).
3. Automate the process of rapidly making novel images.
4. Produce outputs which do not rely on human processes. (2025, 46–47)

Helliwell notes that not all works of AI art will necessarily have all or any of these qualities.

Works without any of these qualities will be decidedly *unmachinic*. Just as there are works of cinema that are uncinematic yet are still works of cinema, there are many works of AI art that do not possess any of the aforementioned artistic properties or make use of the capacities and constraints of the proposed artistic medium of AI art yet still constitute works of AI art. All Helliwell must do is to show that AI art is an artform is to demonstrate that it has the capacity to instantiate artistic properties unique to its medium. Helliwell devotes less of her argument to MSF than to MSV and MSX.

This ultimately leaves Helliwell's list of artistic properties unique to AI art as more of a preliminary suggestion than a definitive, exhaustive list. In the next section I evaluate each of the four unique artistic properties proposed by Helliwell to determine which—if any—are unique to the artistic medium of AI art.

5.5. Evaluating Helliwell's Proposal

With respect to the claim that AI art is uniquely able to synthesize large amounts of data in to a final output, I suggest that this is a difference in degree rather than kind. Consider works of plunderphonics, a genre of music made entirely of samples. It is estimated that the plunderphonics album *Since I Left You* by The Avalanches is constructed of over 3500 samples (Seltmann and Chater 2002). All of these samples are combined to produce a final output (the album). While AI is capable of synthesizing far greater amounts of data, the very act of synthesis is not unique to AI art.

With respect to the claim that AI art is uniquely able to produce images of unreal subjects, I suggest that other artistic media, such as drawing or painting, are also capable of instantiating this artistic property. One could easily make a nonrepresentational painting that does not rely on a real-world subject. Here we can consider Goodman's (1976) symbolic account of pictorial representation to help clarify things. For Goodman, pictures are a symbolic system analogous to other such systems—such as natural language. Similar to how names or predicates pick out particular subjects, Goodman claims that pictures serve an analogous function.

According to Goodman, there are three ways pictures can denote. The first way is individual denotation. A picture of the King of England is an example of individual denotation, as the picture picks out a particular individual. The second way is multiple denotation. A picture

of a woman smiling is an example of multiple denotation, as the picture picks out the class of individuals known as women rather than a particular individual woman. The final way is null denotation. A picture of a unicorn is null denotational, as the picture does not refer to any existing thing. This does not mean that we cannot have pictures of fictional beings such as unicorns. What Goodman is claiming is that such pictures do not represent anything.

Pictures or paintings that are null denotational—such as those of unicorns—do not seem to be meaningfully different from AI-generated images that are unreal. In both cases there is no real-world counterpart that is being represented by the image. Furthermore, an image like Boris Eldagsen’s *The Electrician*, which depicts two women in greyscale, seems to be an example of multiple denotation. While *The Electrician* does not depict any women that exist in the real world, it clearly depicts members of the class of individuals known as women. Finally, it is completely possible for AI to generate images that resemble existing real-world individuals. Imagine a model trained only on images of Elvis Presley. It seems quite likely that such a model would be capable of generating an image that we would recognize as Elvis Presley. This seems to be a clear example of individual denotation, as the image picks out a particular individual.

With respect to the claim that AI art is uniquely able to automate the rapid generation of novel images, I suggest that digital photography can also do this. For example, one could attach a digital camera to a drone programmed to fly along a designated path and change the settings of the camera such that it takes a photo every second. The photos would be novel in the sense that no two would be alike given that the drone’s perspective will have changed over the course of each picture being taken.

While this approach is unusual in digital photography, Helliwell herself states that a particular artistic property need only be capable of being instantiated by the relevant artistic

medium in question in principle. Given that we can clearly imagine a case wherein digital photography is capable of automating the rapid generation of novel images, then this cannot be an artistic property that is unique to AI art.

Finally, with respect to the claim that AI art is uniquely able to produce outputs that do not rely on human processes, I suggest that experimental artforms such as bio-art are capable of this as well. Consider *The Xenotext*, a recent work by experimental poet Christian Bök. The first phase of the project consisted of Bök developing an English language poem that could be translated into DNA nucleotides with the help of a cipher. The DNA would then be translated into RNA, then into a protein, then back into a different English language poem. Bök called the first poem “Orpheus” and the second poem “Eurydice” after the famous characters from Greek mythology.

The second phase of the project involved injecting the “Orpheus”-coded DNA into the bacterium *D. radiodurans*. The bacterium would then convert the DNA into RNA, then into a “Eurydice”-coded protein. The bacterium would then repeat this process indefinitely, creating a never-ending call and response poem. The project was a success and was accompanied by the publication of two books by Bök documenting the experience and the artwork.

The Xenotext is a prime example of an artwork that does not rely on human processes. While Bök developed the poems and the cipher, the artwork proper is the infinite translation of “Orpheus” into “Eurydice.” We might even draw an analogy between Ridler training her own GANs and Bök injecting *D. radiodurans* with the relevant DNA. In both cases the artist prepared the project by “programming” the entity that would produce the artwork. These were the GANs in the case of Ridler’s *Mosaic Virus* and *D. radiodurans* in the case of Bök’s *Xenotext*. While AI

art appears to be capable of producing outputs that do not rely on human processes, this is not an artistic property that is unique to AI art.

It could be the case that AI art is an artistic medium, but not a unique artform. Cross (2024) might give us reason to pause, however, as he provided us with an alternative. Rather than conceiving of AI art as the relevant artistic medium, we might follow Cross's suggestion that AI art can be best understood as a kind of participatory art akin to performance art.

5.6. AI Art as Participatory Art

Cross (2024) provides two different paradigms for understanding AI art. First, there is the production paradigm. This views AI artists as creators who engage with generative AI as either tools or collaborators. Second, there is the exploration paradigm. This views the process of prompting and exploring possible outcomes as a form of participatory art similar to performance art. I provide an overview of both paradigms and then show how Cross's proposed exploration paradigm relates to the medium-specificity thesis.

5.6.1. The Production Paradigm

According to Cross, the most common way of understanding the creation of AI art is through what he terms the production paradigm. This paradigm takes the output of the artistic process to be the site of aesthetic achievement and appreciation. It is the generated image itself that is subject to aesthetic evaluation. In this way, AI can be thought of as a tool or collaborator.

Cross contends that if we are to understand AI as an artistic tool, then we can draw an analogy between the generation of AI images and photography. Similar to the photographer—who makes choices such as how to stage a photograph, which lens and lighting to use, how to develop the film, etc.—the prompt-giver also makes choices insofar as they decide which

prompts to use, how to modify said prompts throughout the artistic process, which generated image or images to select for further use, etc. Both the camera and the AI model can be thought of as tools whose outputs are the site of aesthetic evaluation.

However, one could argue that there is a stark difference between a camera and a diffusion model. The latter involves a degree of stochasticism that is not present in the case of the camera. Thus, while the prompt-giver can guide the image generation process, it is ultimately in some ways beyond their control. Recall that under Young and Terrone's framing, this dynamic recalcitrance is a feature of AI image generation rather than a bug.

Alternatively, if we are to understand AI as a collaborator, then we can draw an analogy between the generation of AI images and the production of a film. Just as a film is produced by a variety of autonomous agents—directors, writers, actors, camera operators, etc.—so too can we understand AI as being an autonomous collaborator in the image production process. Just as actors and writers both contribute towards the creation of a film, so too do prompt-givers and the models to which the prompts are fed.

However, one could argue that there is substantive difference between the collaborators who work together on a film and AI. The latter lacks the capacity to make intentional choices in the same way as the agents responsible for the cocreation of a film. While actors and writers are ultimately limited by things like budget and director discretion, AI is more rigidly constrained by its training data. Even if we grant that AI is capable of making “choices,” it is limited in the kinds of choices it can make in ways that human beings are not.

What both of these conceptualizations—tool and collaborator—have in common is that it is the output of the artistic process that matters with respect to aesthetic evaluation and

appreciation. The production paradigm focuses on the product as the relevant site of aesthetic evaluation and appreciation. In other words, it is the resulting image that is produced that is the artwork.

5.6.2. The Exploration Paradigm

Contra the production paradigm, Cross presents what he terms the exploration paradigm. Here it is not the product of AI image generation that is the site of aesthetic evaluation and appreciation, rather it is the process by which the output is developed. Cross draws an analogy between AI image generation and performance art to make this point.

Cross uses the example of the work of performance artist Maria Abramović to show how we might understand AI image generation as a form of participatory art. Abramović's performance piece *Rhythm 0* involves placing a variety of objects on the floor and providing the audience with instructions to use the items on Abramović in any way that they desire. Cross argues that the performance itself is the site of aesthetic evaluation—not the objects or the instructions.

Similarly, we can think of the process of image generation as belonging to a similar category of art. In the case of AI, the prompt-giver is analogous to the artist/performer, the prompts are analogous to the instructions Abramović provided to her audience, the AI is analogous to an audience participant, and the outputs of the model are analogous to documentations of the performance. In both cases the artwork is the process or performance. However, we can make evaluative judgements about both the process of AI image generation and the performance by viewing the relevant documentation in each case. In the case of the

performance piece, this might take the form of video or photography of the event. In the case of AI image generation this would take the form of observing the AI-generated outputs produced.

Cross argues that a large upshot of this view is that it allows us to understand the *style* of AI art. We might see the *individual style* of a given AI artist come through in their choice of prompts or the strategies they deploy when modifying prompts. Furthermore, we might come to understand the *general style* of AI art by looking to the models and datasets that AI artists are using to create their work.

5.6.3. AI, Participatory Art, and Medium-Specificity

It is easy to apply the medium-specificity thesis to AI art conceived as a form of participatory art. To do so, I must demonstrate two things. First, I must show that evaluations of works of AI art make reference to the medium of participatory art (MSV). Second, I must show that explanations of works of AI art and their artistic properties make reference to the medium of participatory art (MSX). I do not need to show that participatory art instantiates unique artistic properties (MSF), as I am not trying to show that participatory art is a unique artform. If I can show that MSV and MSX both obtain, then we have no immediate reason for favouring Helliwell's view that AI art is a unique artistic medium over Cross's view that AI art belongs to the artistic medium of participatory art.

Recall that for Cross, the prompt-giver is understood as the artist/performer, the AI is understood as the audience/participant, and the process of prompting and generating outputs is understood as a back-and-forth performance. The resulting output—image, text, etc.—is a documentation of the performance and not the artwork proper. The artwork is the performance itself.

First, let us consider evaluations of works of AI art. Both Helliwell and Cross give the example of skillful prompting as being a way of judging the quality of a work of AI art. In Cross's account, the refinement of prompts is done in response to outputs generated by the AI as a kind of back-and-forth performance. The artist or performer (the prompt-giver) reacts to a participant or audience (the AI) in order to produce a work of participatory or performance art (the process of exploration that results in a final output qua documentation). Given that skillful prompting (the choices the artist makes in response to a participant) is part of the performance, our evaluation of prompting as particularly skillful and worthy of praise refers to the artistic medium of participatory art. Ergo, MSV obtains.

Second, let us consider explanations of works of AI art. Here we might turn to the individual style of a given AI artist. While Helliwell could argue that style and artistic properties are in part explained by the limitations of a given model, Cross could respond by saying that we could just as easily explain the style and artistic properties by reference to the choices individual AI artists make when determining which prompts to use. In this case we are not referring to the proposed artistic medium of AI art. Rather, we can explain works of AI art and their artistic properties by referring to the performance given by the prompt-giver to express their individual style of AI art. In this way our explanations of works of AI art and their artistic properties refer to the artistic medium of participatory art. Ergo, MSX obtains.

In the next section I will rescue Helliwell's view by providing a unique artistic property of AI art (the ability to instantiate unique aesthetic properties) that would render it a unique artform. While we could understand AI art as belonging to either the artistic medium of AI art or the artistic medium of participatory art, I will argue that only the former can capture this unique artistic property of AI art.

5.7. A Defence of AI Art

At this juncture we have two competing views. First, we have the view (largely presented by Helliwell) that AI is the vehicle through which the artistic medium of AI art is expressed. In this view evaluations and explanations of works of AI art are made by referencing the set of techniques and practices applied to AI. The artwork is identified as the output produced. Second, we have the view (presented by Cross) that AI is best understood as a participant in the artistic medium (and possibly artform) of participatory art. In this view evaluations and explanations of AI art are made by referencing the performance given by the prompt-giver. The artwork is identified as the process by which the output is produced. The output itself is understood as a documentation of the performance, rather than as the artwork proper.

I will argue that there are at least two aesthetic features that are candidates for being unique to the medium of AI art. It is the ability to produce these aesthetic features that can be understood as an artistic property unique to AI art. I suggest that AI art is the only medium capable of generating works that express what Rini (2024) calls *the artificial sublime*, and an additional aesthetic feature that I call misalignment.

5.7.1. The Artificial Sublime

Rini Draws on Kant's ([1790] 2009) notion of the sublime. For Kant, the sublime is an aesthetic experience whereby first one confronts one's limits as a human and then finds ways one might transcend those limits. Rini gives the examples of the Alps, the Pacific Ocean, and deep space to illustrate Kant's sublime. In each of these cases we are confronted with vast landscapes far greater than our faculties of perception can capture, and yet after the initial period of awe we can come to understand such environments through our reasoning. We can calculate the depth of the

ocean even if we cannot perceive it in its entirety. According to Kant, we can only experience the sublime through encounters with the natural world. For Kant, a painting might evoke an aesthetic experience of the beautiful, but it is only in nature that we might experience the sublime.

Rini extends Kant's notion of the sublime to consider generative AI: "Working with generative AI quickly hints at the vastness of combinatoric space, the unfathomable extent of interesting juxtapositions that no human hand will ever make actual. Our human limits glare at us in comparison to these unending possibilities" (2024, 23). As humans, we simply cannot grasp all of the combinatorial possibilities present when working with generative AI. We are only able to perceive a tiny segment of the full picture, much like in the case of the Pacific Ocean or the Alps. This is the first part of encountering the sublime, wherein we encounter our limits as finite human beings.

The second part, transcending those limits, occurs in the case of generative AI when we come to understand that AI is the product of human cognition. It was humans who figured out how to make such generative models work and it was humans whose data contributed to the training of such models. Thus, according to Rini, AI is an extension of our rational power as humans. Together, these two steps make up the experience of the artificial sublime. Rini holds that the artificial sublime is an aesthetic value unique to generative AI art.

Given that the sublime is typically understood as only being experienced through nature, if we accept Rini's formulation of the artificial sublime then it follows that it is an aesthetic value no other artform is capable of producing.

5.7.2. Misalignment

In chapter 4 I introduced misalignment as an aesthetic feature unique to AI art. Misalignment occurs when there is a significant discordance between the most plausible interpretation a reasonable person would have of a given prompt and the output produced. The weirdness of AI images identified by Helliwell (2023) is one such example of misalignment.

For example, the most plausible interpretation a reasonable person would have of the prompt “orange cat” would be of an ordinary orange cat with four legs and a tail. However, many models—especially earlier ones—would often interpret this prompt in bizarre ways, such as generating orange amorphous blobs that vaguely resemble cats. This is a case of the aesthetic feature of misalignment.

Some artists have been using AI to generate images that possess misalignment as a way of exposing algorithmic bias. I call this style of AI art bias-prompting. AI artists have been engaging in bias-prompting to perform institutional critique (Ramsden 1975). While institutional critique is not unique to AI art, the mechanism by which the critique is enacted (by creating images that possess misalignment) seems to be unique to the medium of AI art.

Art often subverts our expectations. We might think that the abstract nature of Cubism has a lot in common with the weirdness of AI art. However, Cubist works are intended to look a certain way. Misalignment often occurs when our intentions and expectations differ from the outputs generated by AI. Even in styles wherein there is a degree of dynamic recalcitrance, such as action painting or aleatoric music, we still have our expectations met. For example, even if I do not know how the paint will settle when doing action painting, I know that there will be paint on a canvas. Likewise, with aleatoric music, I do not know which note will be decided upon prior to rolling a die. However, I do know that it will be a note in the chromatic scale.

The difference with misalignment is that the discordance between how a reasonable person would plausibly interpret a given prompt and the output that the AI produces is significant. This is why we often interpret instances of misalignment as mistakes on the part of the AI. We understand Cubist works as being abstract due to creative choices made by the artist. However, in the case of AI, the discordance is so significant and creative intentions seem absent, so we take instances of misalignment to be mistakes in need of correction.

5.8. AI Art as Artform

Recall that there are two positions on the table. The first holds that AI art is a unique artistic medium as well as a unique artform. The second holds that AI art can be best understood as a form of participatory art. I will now give a defence of the first position.

There are two reasons why we ought to reject Cross's suggestion of conceiving of AI art as a kind of participatory art. The first reason is that the ability to produce works with the aesthetic features of the artificial sublime or misalignment is an artistic property that seems unique to AI art (MSF). Even if there are works of AI art that do not instantiate these aesthetic properties, we still ought to understand these works as belonging to the artform of AI art. Recall that *My Dinner With André* is a work of cinema that is uncinematic and more closely resembles a play. However, we do not say that it is a work in the artform of theater, as it was recorded on film and projected on cinema screens and is thus a work of cinema. Likewise, we can say that a work of AI art that is unmachinic is still a work in the artform of AI art as it is a work in the medium of AI art.

The fact that AI art has the capacity to produce unique aesthetic features is what makes it an artform. Understanding AI art as being participatory art cannot explain this capacity as

participatory art is incapable of instantiating the aesthetic features of misalignment and the artificial sublime. This is because the capacity to produce misalignment and the artificial sublime is unique to AI art.

The second reason for rejecting Cross is that it seems disanalogous to compare the generation of AI art to a performance. In the case of AI, there is no audience outside of the AI. Video games involve computers reacting to player choices and yet we do not consider the playing of video games as engaging in participatory or performance art. It is a stretch to say that an individual prompting a computer constitutes a performance in the absence of an audience.

In summary, AI can be understood as the vehicle through which the artistic medium of AI art is expressed. AI art can additionally be understood as an artform if we accept that it possesses the unique artistic property of being capable of producing works that express misalignment and the artificial sublime. This view is preferable to the alternative provided by Cross that views AI as a participant and AI art as a kind of participatory art, as it seems untenable to say that prompting AI is like a performance.

In the next chapter I return to the question of adjudicating authorship by exploring the case of AI-generated texts.

Chapter 6: Texts Without Authors

6.1. Ascribing Literary Meaning in the Case of AI

While the question of who or what fixes the meaning of literary texts is not new, developments in Large Language Models (LLMs) both complicate and illuminate the question. If we understand some AI-generated works as *texts*, and such texts have *meaning*, then there is a puzzle in determining who or what fixes said meaning. Views that defer to authorial intent are of little help, given the absence of a clear *author* in these cases. I argue that such texts can and do acquire meaning irrespective of the presence of an author.

This chapter proceeds as follows. First, I provide a brief overview of how LLMs—such as GPT-4—work on a technical level. I then draw a tripartite distinction between *Illusory Texts*, *Artificial Texts*, and *Traditional Texts*. I argue that some Artificial Texts have fixed meanings. Then, I give an overview of six popular views on how the meaning of texts are fixed. I demonstrate that only two of the views, *conventionalism* and *postulated author hypothetical intentionalism*, show any promise of solving the puzzle. I argue against conventionalism and in favour of postulated author hypothetical intentionalism. I conclude by considering some implications for our understandings of texts generated by AI.

6.2. Large Language Models

Contemporary LLMs are a kind of *generative AI* that produce text outputs based on user inputs in the form of *prompts*. Contemporary LLMs are *Deep Learning Models* which are *neural networks* with many layers. A neural network is an interconnected structure of nodes—called *neural units*—that process input and generate output. LLMs are trained on large *datasets* of text scraped from the internet, taken from books, etc. Words are then encoded into *word embeddings*,

vector representations of words in a multidimensional space, with related words clustering together (Stinson 2024). For example, if one were to input the prompt “What is the capital of Canada?” the model would recognize “capital” and “Canada” as existing syntactically in relation to one another and would likely generate “Ottawa” as output.

Many contemporary LLMs use a model architecture called a *Transformer* (Radford et al. 2018, 2019; Brown et al. 2020; Ouyang et al. 2022). Popularized by Vaswani et al (2017) the Transformer architecture is a neural network wherein each unit uses a mechanism called *self-attention*, which is capable of “relating different positions of a single sequence in order to compute a representation of the sequence” (2). Rather than analyse the elements of a sentence in a linear fashion, LLMs built using Transformer architecture are able to analyse the different elements of a sentence nonlinearly. This provides stronger results in text generation, as more relationships between word embeddings can be considered concurrently, rather than processed one at a time.

For example, if one were to ask an application like ChatGPT “What does a cat like to eat?” a simple response would be “food” as the words “eat,” and “food” are often found in close proximity to each other in many English language sentences. However, the additional context of “cat” and “likes” might provide a more sophisticated response, such as “fish,” or “tuna.” LLMs are thus able to generate sentences based on both the likelihood of certain words being in close proximity to one another in the training data and the likelihood of certain word embeddings being clustered together. This is why prompts such as “Write me a story about a knight slaying a dragon in the style of Edgar Allen Poe” often yield satisfying results. The model is able to recognize both Poe’s vocabulary and syntax, as well as the tropes surrounding fantasy stories about knights slaying dragons.

Given that LLMs work primarily by pattern recognition, it is not as though they have *intentional attitudes* toward the text that they generate.⁵ As such, they cannot be properly considered *authors* in the sense of being the intentional meaning-makers of texts. Likewise, the prompt-giver could not be considered an author proper, as they do not generate any of the text, nor do they know in advance what the content of the text output will be.

There are two possible moves one could make at this juncture. The first move would be to deny that such outputs can be considered texts proper, possessing neither an author nor literary meaning. The second move would be to accept the outputs as texts, but to deny that they have a fixed literary meaning. The former move would come from thinkers following Gracia (1995, 1996), whereas the latter might come from a follower of Barthes ([1967] 1977). In the next two sections, I provide an argument against each in turn, ultimately endorsing the view that AI-generated works can be texts, and that such texts have fixed meanings.

6.3. Three Kinds of Texts

One way to solve the problem of ascribing meaning to AI-generated works is to deny that such works can be literary texts.⁶ There are several reasons why we should not take this to be the case.

We might consider works like David Jhave Johnston's *ReRites*, a collection of AI-generated poetry. *ReRites* is a twelve volume poetry set compiled between 2017 and 2018 consisting of AI-generated output arranged and edited by Johnston and later released in an edited format alongside critical essays (2019). Elsewhere, Allison Parish's *Articulations* (2018) takes a similar approach by training an LLM on a dataset of public domain poetry and publishing a

⁵ This claim about AI in general has been defended by others, perhaps most notably by John Searle (1980) and Ned Block (1981). I will defer to previous scholarship and not provide my own defense of the claim that AI lacks intentionality.

⁶ Followers of Gracia (1995, 1996) would likely take this approach.

curated set of results. Works such as these—and *electronic literature* more broadly—have been taken up in academic circles among literature scholars as both subjects of criticism (MacBeth 2023) and texts that challenge traditional understandings of literature (Hayles 2018).

ReRites and *Articulations* were presented as literary works and taken up as such. This is in line with institutional theories of art such as those by Danto (1964) and Dickie (1969, 1974) that hold for a given work to be a work of art, it must be presented to an *artworld*. Levinson (1996) holds a similar view about literary works. According to institutional theories of art, a work like *ReRites* would indeed be a work of literature. We might also consider the move in contemporary literature to consider any work a potential text (Goldsmith 2011). According to conceptual poets, even a phonebook can be a literary text if it is presented as such.

This line of argumentation would likely not persuade those in Gracia's (1995, 1996) camp. For Gracia, a text must contain an intentional element. In order for a work to constitute a text proper, it must be the product of an intentional agent. Shakespeare's *King Lear* is a text, a cluster of rocks randomly arranged by the wind to spell out a verse of Shakespeare is not. However, something strange occurs when an artist takes a photo of the rocks, labels them "Shakespeare Rocks" and presents the work to an artworld.⁷ If the work is taken up by audiences and critics, it would have a claim to text status.⁸

There seems then, to be a distinction between *King Lear*, the cluster of rocks, and "Shakespeare Rocks." We might draw the lines this way: *King Lear* is a *Traditional Text*. It is the product of an intentional agent—an author—who decided to create the work. The cluster of rocks is an *Illusory Text*. In other words, while it might have the appearance of a text, it is no text at all.

⁷ Here I am playing with Walton's (1990) famous "Cracks in a Rock" thought experiment.

⁸ Likewise with other found texts discussed by scholars like Goldsmith (2011).

“Shakespeare Rocks” appears to occupy a strange middle ground. While it was not created by an intentional agent, its status as a literary work depends on an intentional agent. The same goes for AI-generated works. They cannot be said to have been created by an intentional agent, but their creation is dependant on the agency of a curator. In this sense we might think of works like “Shakespeare Rocks” and *ReRites* as *Artificial Texts*—texts without authors.

One could argue that in the case of “Shakespeare Rocks” the photographer is the author of the work, given that we consider photographers to be the authors of the pictures they take in general. Here I make the distinction between “Shakespeare Rocks” qua photograph and “Shakespeare Rocks” qua text. While it is true that the photographer is the author of “Shakespeare Rocks” qua photograph, they are not the author of “Shakespeare Rocks” qua text. Here it is useful to consider Christy Mag Uidhir’s (2011) notion of x being the author of a work W as an F. Consider the example of films. The writer of a film is the author of the film as a script, whereas the director of a film is the author of the film as a film. Similarly, the photographer is the author of “Shakespeare Rocks” as a photograph, but not as a text.

Furthermore, the photographer did not imbue the text with meaning, they noticed something interesting about the rock formation (that it spells out a verse of Shakespeare) and presented it to an audience for consideration. The photographer acts as a curator in this scenario rather than an author. Cases like *Articulations* are similar, insofar as the text is produced by AI and then taken up by a curator and presented to an audience. “Shakespeare Rocks,” *Articulations*, and *ReRites* possess interesting textual meaning irrespective of a curator. However, it is the curator that transforms the text from an Illusory Text to an Artificial Text.⁹ The curators

⁹ It might be the case that all Artificial Texts begin their existence as Illusory Texts.

in these examples are inviting the audience to consider these works as texts, not as photographs or other artforms. Left alone, “Shakespeare Rocks” would remain an Illusory Text. However, it becomes an Artificial Text when a curator presents it as such. The role of the curator differs from the role of the author. In this way we can consider Artificial Texts to be texts without authors.

In summary, we can think of three distinct kinds of text. First, there are Illusory Texts, which take on the appearance of texts but are not texts themselves. Then, there are Traditional Texts, which are the products of intentional agents. Finally, there are Artificial Texts, which are not the products of intentional agents, but whose status as a text depends on the agency of a curator. AI-generated texts are thus a different kind of text than Traditional Texts, yet we can think of them as texts, nonetheless.

However, despite their status as texts, it could still be the case that Artificial Texts—such as AI-generated works or found poetry—do not have fixed meanings. I address this position in the next section.

6.4. Meaning and Artificial Texts

In this section I pursue the question as to whether some Artificial Texts have fixed meanings. Before addressing this, however, I will first define *literary meaning* and offer three theories for understanding literary meaning.

I take literary meaning to be distinct from semantic meaning insofar as a work of literature can come to mean more than the sum of its atomic parts. For example, a literal reading of Orwell’s *Animal Farm* reveals a story about animals on a farm. However, a literary reading of the same text reveals an allegorical work about political struggle. The semantic meaning is the former, whereas the literary meaning is the latter. According to Olsen, “...it has become a critical

commonplace that the literary work is a verbal expression, a verbal construct or an utterance, and that as such its peculiar nature is defined through the special *way in which it means*” (1982, 13). On this account, literary meaning is not reducible to semantic meaning, though semantic theories inform literary ones.

Olsen (1982) identifies three kinds of theory of literary meaning: there are *autonomy theories*, *semiotic theories*, and *intentionalist theories*. Autonomy theories (Frye 1957; Beardsley 1958; Richards 1965) hold that the meaning of a text is immanent to the text. Semiotic theories (Propp 1968; Culler 1981) hold that meaning is determined through the interpretation of the relationship between the signs and systems that govern a given text.¹⁰ Intentionalist theories (Hirsch 1967; Juhl 1980) hold that the meaning of a text is imbued by the intention of the author of the text.¹¹ For both autonomous and intentionalist theories, a given text has a fixed literary meaning. Semiotic theories differ as to whether they consider texts to have fixed meanings. Of those that do not hold that texts have fixed meanings, Barthes ([1967] 1977) is most notable.

For Barthes, the author of a text does not hold a privileged position with respect to the meaning-making of a text. Rather, the meaning of a text is determined collectively by the audience—of whom the author is a part. This can be read as conservatively as there being one meaning generated collectively by the audience, or as radically as there exists a different meaning for every reader. It might be more accurate in the second case to say that the text does not possess a meaning at all, rather there are only audience interpretations.

¹⁰ Texts can either be *closed*, such that no outside understandings of signs are permissible, or *open*, wherein the signs and systems that make up a text are in conversation with other discourses.

¹¹ I discuss these in greater detail in the next section.

There is a major problem with this view however, as we can be mistaken about what a text means. In light of this, some interpretations of a text are more correct than others. Consider *Animal Farm*. It would be incorrect to interpret the text as being an allegory for the 2016 US election, as the writing and publication of *Animal Farm* greatly predates the election.¹² One could argue that over time texts acquire new meanings. However, this ignores the relevance of the context in which a text was produced and generates bizarre ahistorical readings.

It is more reasonable to interpret *Godzilla* as a metaphor for nuclear weapons than it is to interpret the H.G. Wells' *War of the Worlds* as one. This is because nuclear weapons did not exist at the time Wells was active and publishing. While we can *interpret* texts in a variety of ways, that does not entail that such interpretations correspond to *meanings*. If the goal of interpretation is to discern the meaning of a text, then not all interpretations of a text are good interpretations as they do not fulfill this function.¹³ It does not follow from this that all Artificial Texts have fixed meanings. It is very likely that many do not.¹⁴

According to literary critic Gérard Genette (1992) *transtextuality* is everything that brings a text into relation, either explicitly or implicitly with other texts. Under the umbrella of transtextuality is *metatextuality*, that which connects a commentary of a text to the text itself. Artificial Texts are often conceptual in nature. As such, much of the meaning one is to find in such texts exists through their meta level commentary. This can be commentary on other texts, or

¹² It remains an open question as to *when* and *how* a given text acquires its meaning. I discuss this issue in the next section.

¹³ It is also possible, in principle, for a text to have multiple fixed meanings.

¹⁴ This is not unique to Artificial Texts. There are Traditional Texts that appear to lack fixed meaning, such as James Joyce's *Finnegans Wake*, a "novel" consisting only of surreal stream of consciousness word play and devoid of any obvious plot or characters.

on the text in question. In what follows I give two examples of Artificial Texts that have fixed meanings at the metatextual level.

First, consider *Via: 48 Dante Variations* by Caroline Bergvall. As a work of found poetry, *Via* is a compilation of 48 iterations of the first canto of Dante's *Inferno*. It begins with the original Italian and follows with 47 unique English translations. The work explores the process of translation and what is lost and gained through the act. A direct reading of any individual part of the text does not reveal its meaning, however, by reading the text as a whole—with the various translations juxtaposed—one can come to understand the work as being about translation.

Second, consider *Death of an Author* by Stephen Marche. *Death of an Author* is a detective novella written almost entirely with the use of models like ChatGPT. The story follows literary critic Gus Dupin as he investigates the murder of celebrated novelist Peggy Firmin. Dupin becomes embroiled in a twisted plot involving the LLM company Marlow AI as he himself becomes a murder suspect. While on the surface a work of detective fiction, the novel explores the relationship between humans and AI, especially as they pertain to the realm of literature. In that sense the novella is a mirror of its own creation and acts as a metaphor for its own process.

Both *Via* and *Death of an Author* are Artificial Texts with what appear to be fixed meanings, with the former being about translation and the latter about human-AI relations. Granted that both texts have fixed meanings, then the puzzle of who or what fixes the meaning of said texts becomes apparent. In the next section I outline six competing views on attributing literary meaning, and jettison four as viable contenders.

6.5. Six Views of Ascribing Literary Meaning

According to Irvin (2006), there are six leading theories of how a literary text's meaning is fixed.¹⁵ These are extreme actual intentionalism (EAI), two versions of modest actual intentionalism (MAI1 and MAI2), conventionalism, actual author hypothetical intentionalism (AAHI) and postulated author hypothetical intentionalism (PAHI).¹⁶ I will go over each in brief detail before dismissing EAI, MAI1, MAI2, and AAHI as viable solution bearers.

Extreme actual intentionalism (Irwin 1999, 2015; Stock 2017; García-Carpintero 2019, 2023) is the view that it is the author's intent that fixes the meaning of a text, irrespective of linguistic conventions. This means that if an author intended for the sentence "the apple is red" to mean "the banana is yellow," then we ought to take the second to be the meaning of the first in that context. Modest actual intentionalism (Carroll 2000) deals with this strange feature of EAI by taking the view that when an author's intent coincides with linguistic convention, it is the author's intent that fixes the meaning. Proponents of MAI1 hold that when the author's intent diverges from linguistic convention, the affected portion of the text is meaningless. Proponents of MAI2 hold that when this occurs, either meaning is fixed by linguistic convention, or that portion of the text remains ambiguous (Stecker 2003). MAI2 can collapse into conventionalism if a given text is especially ambiguous.

Conventionalism holds that it is linguistic convention that fixes the meaning of a text (Beardsley 1958), although some have argued that context ought to be considered as well (S.

¹⁵ All of the theories presented rely on an intentionalist account of meaning, except for conventionalism, which relies on an autonomous account of meaning.

¹⁶ I follow Irvin (2006) and Nehamas (1981) in labelling this author the *postulated author*. Other names have been suggested, such as the *apparent artist* (K. Walton 2008), the *fictional author* (Currie 1990), and the *hypothetical author* (Lin 2023). I take all of these terms to be more or less equivalent for my purposes.

Davies 2005). Conventionalists disagree over what to do in cases where the text is ambiguous, but do not defer to an author in such cases. Hypothetical intentionalism comes in two forms. Actual author hypothetical intentionalism holds that the meaning of a given work is fixed by the best hypotheses an informed audience would form about the actual author's intentions (Levinson 1996). Postulated author hypothetical intentionalism differs insofar as it is not the actual author of a work, but an ideal author that possesses additional features such as full mastery of language (Lin 2018, 2023).¹⁷ AAHI often collapses into a version of actual intentionalism, while PAHI runs the risk of collapsing into conventionalism. However, as I will later argue, there are cases where the distinction between conventionalism and PAHI becomes meaningful such that we might prefer PAHI.

In the case of AI-generated works, it does not make sense to defer to an author, as Artificial Texts are *texts without authors*. This fact rules out EAI, MAI1, MAI2, and AAHI as contenders for a theory of ascribing meaning that can solve the puzzle of AI-generated texts, given that they all require the presence of an author for meaning to be fixed in at least some cases. For EAI and AAHI, there is simply no author to whom one could look to for meaning. In MAI1, the AI-generated text in question would have to be conceived as meaningless—I have already given cases where we should not take this to be true. In MAI2, the meaning of the text would be fixed by linguistic convention, which is to say that in cases of AI-generated works, MAI2 *always* collapses into conventionalism.

¹⁷ One might wonder about the difference between PAHI and death of the author (Barthes [1967] 1977). While in both cases it is the audience that fixes the meaning of a text, the postulated author of PAHI serves as a limiting function on the number and kinds of possible interpretations. If one agrees with Barthes about the death of the author, then one could plausibly argue that it is merely the audience that fixes the meaning of a text. As I hold that some interpretations of texts are more correct than others, I do not endorse this view.

There is one additional move that defenders of EAI could make in an attempt to salvage the explanatory power of EAI in cases wherein the text in question is AI-generated. Irwin (2015) draws a distinction between *meaning* and *significance*. The former is the interpretation of a text intended by its author. The latter is any interpretation of a text that differs from that intended by its author. Queer readings of literature that lacks explicit reference to queerness would count as examples of significance. On this account it follows that in the absence of an author, AI-generated text has no fixed meaning but many legible readings with significance.

I have already argued that some Artificial Texts have fixed meanings, however one need not turn to AI-generated text for examples of texts without fixed meanings. Consider the song “I Am the Walrus” by *The Beatles*. The song was written without the explicit intent of having a meaning, and yet music critics and audiences alike have drawn their own salient interpretations of the song’s nonsense lyrics. We might also consider the art movement dadaism as a way of making texts without fixed meaning that still possess significance. However, there is a huge leap from the nonsense lyrics of “I Am the Walrus” or dadaist sound poetry to AI-generated text. While the former two seem to be examples of intentional meaninglessness, the latter appears to be the opposite: unintentional meaningfulness. Both “I Am the Walrus” and dadaist poems feature odd word choice and deeply unusual syntax and semantics.

Popular LLMs like ChatGPT seem to be doing something different insofar as the syntax and semantics of AI-generated sentences utilize linguistic conventions to produce meaningful content. There is a stark difference between the nonsense phrase “goo goo g’joob” and a work such as *Death of an Author*. The former lacks obvious meaning, whereas the latter contains obvious meaning. Reducing the meaning of a text like *Death of an Author* to significance seems to be a case of linguistic confusion. If there is only one significance of a text, it would make

more sense to consider that the meaning of said text, rather than to maintain that the text in question has no meaning. In light of this, EAI remains inappropriate as a candidate for explaining meaning in cases of AI-generated texts.

While I have ruled out four of the six competing theories of ascribing meaning to a text, conventionalism and PAHI remain on the table. In the next section I argue that we ought to reject conventionalism—despite its seeming potential to resolve the puzzle—and instead embrace PAHI

6.6. Fixing the Meaning of Artificial Texts

While at first glance conventionalism seems as though it can adequately solve the puzzle of attributing meaning to AI-generated texts—by holding that it is *only* linguistic convention that fixes meaning, thus circumnavigating the problem of the absent author—it runs into problems insofar as it is unable to account for meaning found at a metatextual level.

Consider the two Artificial Texts discussed earlier, *Via* and *Death off an Author*. There is nothing about the individual words or passages of *Via* that would lead to the interpretation that the text is about translation. Rather, the text must be taken holistically. It is only by understanding *Via* as being constructed by juxtaposing different translations of Dante can we understand it as a work about translation. *Death of an Author*, on the other hand, *can* be interpreted through the individual parts of the text. However, the larger metatextual point of the work itself being imbricated in its meaning—it is an exemplar of the themes of the narrative—is lost.

Thus far I have identified two main problems in ascribing meaning in the case of AI-generated texts. The first comes with most forms of intentionalism that rely on the presence of an author for a text to acquire meaning. In the case of AI, there is no author to whom we can defer. The second problem lies with conventionalism's inability to account for meaning on a metatextual level. I will show how PAHI is able to effectively deal with both of these problems.

First, the problem of authorship. According to PAHI, it is not the author who fixes the meaning of a text. Rather, it is the audience's best hypotheses, considering relevant evidence, of what meaning an ideal author—who possesses such features as mastery of the English language—would imbue within the text that fixes meaning. As it is the audience that fixes meaning with their informed hypotheses given the supposition of an ideal author, the problem of an absent author is circumnavigated entirely. We posit an author—call them Stephen Marche — that possesses mastery of style and language in addition to having reasonable intentions surrounding the meaning of their work. In this way, the postulated author Stephen Marche serves a role in fixing the meaning of AI-generated text in the absence of an actual author Stephen Marche.

Second, given the postulation of an ideal author, PAHI can account for metatextual meaning in ways conventionalism cannot. The postulated author of PAHI possesses mastery of language, including the use of metatextual writing. Rather than understand the meaning of a text to be merely the sum of individual components, PAHI can consider texts holistically to unlock their meanings at the metatextual level.

While both conventionalism and PAHI can deal with the first problem—the absence of an author, only PAHI can deal with the second—how texts can come to have meaning at a

metatextual level. Given this, we ought to accept PAHI as being able to account for the fixing of meaning in AI-generated texts.

6.7. Consequences

In this chapter I have argued three things. First, that AI-generated texts ought to be considered Artificial Texts—texts without authors. Second, that Artificial Texts sometimes have a fixed literary meaning. Third, the view that best explains how the literary meaning of Artificial Texts is fixed is Postulated Author Hypothetical Intentionalism.

There are two main consequences of my view. The first is that we now have the language of Artificial Texts to understand AI-generated works as texts proper. As Artificial Texts are distinct from Traditional Texts, we know that the same rules and implications need not apply to both kinds of text. This distinction also helps explain how works such as found texts can be considered texts proper.

The second consequence is that once we determine an Artificial Text has a fixed meaning, we have an interpretive framework—PAHI—for discerning said meaning. The difficulty now lies in constructing a rubric or heuristic for determining which Artificial Texts have fixed literary meanings, and which Artificial Texts may be (re)interpreted by the audience ad nauseum.

Chapter 7: Towards a New Philosophy of AI Art

In this dissertation, I have articulated the beginnings of a new philosophy of AI art. Here I have defended five main claims.

First, I have argued that according to contemporary accounts of art, AI-generated images can sometimes count as artworks.

Second, I have argued that the class of AI artwork that Nannicelli identifies as mass AI-art largely counts bad art—according to a failed organic whole view of bad art.

Third, I have argued that there exists a unique style of AI art that I term bias-prompting. This style exploits the presence of algorithmic bias within diffusion models to produce images that illustrate how such models (re)produce and (re)inscribe oppressive racist and sexist stereotypes. I contend that this style has the capacity to produce images with a unique aesthetic feature that I call misalignment.

Fourth, I have argued contra Young and Terrone that AI does not on its own constitute a novel artistic medium. However, AI is the vehicle through which the artistic medium of AI art is expressed.

Finally, I have argued that literary texts produced by LLMs are Artificial Texts without authors. Instead, such texts have a curator. The best way of understanding how Artificial Texts acquire literary meaning is postulated author hypothetical intentionalism.

While I have provided some starting points for considering questions about aesthetics and authorship in the case of AI art, there remain pressing ethical questions surrounding the practice.

For example, it remains an open question if synthetic images constitute forgery (Helliwell 2023), plagiarism (Wojtkiewicz 2023; Marcus and Southen 2024), or theft (Goetze 2024). One promising avenue for resolving these issues presents itself in the literature on the autographic/allographic distinction (Goodman 1976; Zeimbekis 2012; D’cruz and Magnus 2014). Of particular interest is the work of Terrone (2018) who, drawing on the work of Davies (2010) gives rigorous conditions under which forgery and plagiarism obtain.

Furthermore, we might wonder whether there are aesthetic goods worth creating and preserving even if it turns out to be the case that AI art constitutes theft. Thus, we might see a tension between certain ethical systems that would prohibit theft (such as deontology) and certain aesthetic systems that would encourage the production of aesthetic goods (such as virtue aesthetics).

This is of course a non-exhaustive list of future research programmes for philosophy of AI art. Further questions surrounding how we can best appreciate such works, how we ought to evaluate them, and the role of creativity in AI art are also relevant.

This dissertation serves to contribute to the growing body of literature on the topic of AI art. While AI technologies might change and the AI bubble might burst, artists will continue to find new and interesting ways to engage with emerging technologies in their artmaking practices.

References

Abell, Catharine. 2009. "Canny Resemblance." *The Philosophical Review* 118 (2): 183–223.

<https://doi.org/10.1215/00318108-2008-041>.

Agarwal, Sandhini, Gretchen Krueger, Jack Clark, Alec Radford, Jong Wook Kim, and Miles Brundage. 2021. "Evaluating CLIP: Towards Characterization of Broader Capabilities and Downstream Implications." arXiv:2108.02818. Preprint, arXiv, August 5.

<https://doi.org/10.48550/arXiv.2108.02818>.

Algander, Per. 2024. "What's so Good about Bad Art?" *Philosophy Compass* (Ithaca) 19 (11):

n/a. <https://doi.org/10.1111/phc3.70015>.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2022. "Machine Bias *." In

Ethics of Data and Analytics. Auerbach Publications.

Anscomb, Claire. 2025. "Who Authors AI Art? (And Why Does It Matter?)." *AI & SOCIETY*,

ahead of print, April 15. <https://doi.org/10.1007/s00146-025-02325-7>.

Barthes, Roland. (1967) 1977. "The Death of the Author." In *Image, Music, Text*, translated by

Stephen Heath. Fontana.

Beardsley, Monroe C. 1958. *Aesthetics: Problems in the Philosophy of Criticism*. Harcourt,

Brace.

Bianchi, Federico, Pratyusha Kalluri, Esin Durmus, et al. 2023. "Easily Accessible Text-to-Image

Generation Amplifies Demographic Stereotypes at Large Scale." *Proceedings of the 2023*

- ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA), FAccT '23, June 12, 1493–504. <https://doi.org/10.1145/3593013.3594095>.
- Blau, Yochai, and Tomer Michaeli. 2018. “The Perception-Distortion Tradeoff.” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June, 6228–37. <https://doi.org/10.1109/CVPR.2018.00652>.
- Block, Ned. 1981. “Psychologism and Behaviorism.” *The Philosophical Review* (Boston) 90 (1): 5–43. <https://doi.org/10.2307/2184371>.
- Boden, Margaret A., and Ernest A. Edmonds. 2009. “What Is Generative Art?” *Digital Creativity* 20 (1–2): 21–46.
- Brady, Emily. 2003. *Aesthetics of the Natural Environment*. University Press.
- Brady, Emily. 2011. “The Ugly Truth: Negative Aesthetics and Environment.” *Royal Institute of Philosophy Supplement* 69: 83–99.
- Brown, Tom B., Benjamin Mann, Nick Ryder, et al. 2020. “Language Models Are Few-Shot Learners.” arXiv:2005.14165. Preprint, arXiv, July 22. <https://doi.org/10.48550/arXiv.2005.14165>.
- Carroll, Noël. 1992. “Art, Intention, and Conversation.” In *Intention and Interpretation*, edited by Gary Iseminger. Temple University Press.
- Carroll, Noël. 1998. *A Philosophy of Mass Art*. Clarendon Press.
- Carroll, Noël. 1999. “Horror and Humor.” *The Journal of Aesthetics and Art Criticism* 57 (2): 145–60. <https://doi.org/10.2307/432309>.

- Carroll, Noël. 2000. "Interpretation and Intention: The Debate Between Hypothetical and Actual Intentionalism." *Metaphilosophy* 31 (1/2): 75–95.
- Carroll, Noël. 2003. "Forget the Medium!" In *Engaging the Moving Image*. Yale University Press.
- Cascales, Raquel. 2023. "Interpreting AI-Generated Art: Arthur Danto's Perspective on Intention, Authorship, and Creative Traditions in the Age of Artificial Intelligence." *The Polish Journal of Aesthetics*, no. 71: 17–29.
- Charles, David R. 2025. "AI as Artist: Agency and the Moral Rights of Creative Works." *AI and Ethics*, ahead of print, February 19. <https://doi.org/10.1007/s43681-025-00679-8>.
- Chen, Muxi, Yi Liu, Jian Yi, et al. 2024. "Evaluating Text-to-Image Generative Models: An Empirical Study on Human Image Synthesis." arXiv:2403.05125. Preprint, arXiv, March 8. <https://doi.org/10.48550/arXiv.2403.05125>.
- Chiang, Ted. 2024. "Why A.I. Isn't Going to Make Art." The Weekend Essay. *The New Yorker*, August 31. <https://www.newyorker.com/culture/the-weekend-essay/why-ai-isnt-going-to-make-art>.
- Cooper, David E. 2003. "In Praise of Gardens." *The British Journal of Aesthetics* 43 (2): 101–13. <https://doi.org/10.1093/bjaesthetics/43.2.101>.
- Cross, Anthony. 2024. "Tool, Collaborator, or Participant: AI and Artistic Agency." *The British Journal of Aesthetics*, December 21, ayae055. <https://doi.org/10.1093/aesthj/ayae055>.

- Culler, Jonathan D. 1981. *The Pursuit of Signs: Semiotics, Literature, Deconstruction*. Cornell University Press.
- Currie, Gregory. 1990. *The Nature of Fiction*. Cambridge University Press.
- Danks, David, and Alex John London. 2017. "Algorithmic Bias in Autonomous Systems." *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (Melbourne, Australia), IJCAI'17, August 19, 4691–97.
- Danto, Arthur. 1964. "The Artworld." *The Journal of Philosophy* 61 (19): 571–84.
- Danto, Arthur. 2003. *The Abuse of Beauty: Aesthetics and the Concept of Art*. Open Court.
- Davies, David. 2009. "Medium in Art." In *The Oxford Handbook of Aesthetics*, edited by Jerrold Levinson. Oxford Academic.
- Davies, David. 2010. "Multiple Instances and Multiple 'Instances.'" *The British Journal of Aesthetics* (Oxford) 50 (4): 411–26. <https://doi.org/10.1093/aesthj/ayq030>.
- Davies, Stephen. 2005. "Beardsley and the Autonomy of the Work of Art." *The Journal of Aesthetics and Art Criticism* 63 (2): 179–83.
- D'cruz, Jason, and P. D. Magnus. 2014. "Are Digital Images Allographic?" *The Journal of Aesthetics and Art Criticism* 72 (4): 417–27. <https://doi.org/10.1111/jaac.12117>.
- Dickie, G. 2000. "Art and Value." *The British Journal of Aesthetics* 40 (2): 228–41. <https://doi.org/10.1093/bjaesthetics/40.2.228>.
- Dickie, George. 1969. "Defining Art." *American Philosophical Quarterly* 6 (3): 253–56.

- Dickie, George. 1974. *Art and the Aesthetic : An Institutional Analysis*. Cornell University Press.
- Du, Yifan, Zikang Liu, Junyi Li, and Wayne Xin Zhao. 2022. “A Survey of Vision-Language Pre-Trained Models.” arXiv:2202.10936. Preprint, arXiv, July 15.
<https://doi.org/10.48550/arXiv.2202.10936>.
- Duchamp, Marcel. (1957) 1975. “The Creative Act.” In *The Essential Writings of Marcel Duchamp*, edited by Michel Sanouillet and Elmer Peterson. Thames and Hudson.
- Dyck, John, and Matt Johnson. 2017. “Appreciating Bad Art.” *The Journal of Value Inquiry* (Dordrecht) 51 (2): 279–92. <https://doi.org/10.1007/s10790-016-9569-2>.
- Ensslin, Astrid. 2012. “Computer Gaming.” In *The Routledge Companion to Experimental Literature*, edited by Joe Bray, Alison Gibbons, and Brian McHale. Routledge.
- Fazelpour, Sina, and David Danks. 2021. “Algorithmic Bias: Senses, Sources, Solutions.” *Philosophy Compass* (Ithaca) 16 (8): n/a. <https://doi.org/10.1111/phc3.12760>.
- Feral File. n.d. “Experience Minne Atairu’s Blonde Braids Study II (2023) on Feral File.” Accessed November 28, 2025. <https://feralfile.com>.
- Foucault, Michel. (1969) 1998. “What Is an Author?” In *Aesthetics, Method, and Epistemology*, edited by James D. Faubion, translated by Robert Hurley. Essential Works of Foucault, 1954–1984, edited by Paul Rabinow. The New Press.
- Frye, Northrop. 1957. *Anatomy of Criticism: Four Essays*. Princeton University Press.
- García-Carpintero, Manuel. 2019. “On the Nature of Fiction-Making: Austin or Grice?” *The British Journal of Aesthetics* 59 (2): 203–10. <https://doi.org/10.1093/aesthj/ayy054>.

- García-Carpintero, Manuel. 2023. "The Semantics of Fiction." *Mind & Language* 38 (2): 604–18. <https://doi.org/10.1111/mila.12412>.
- Gaut, Berys. 2010. "The Role of the Medium." In *A Philosophy of Cinematic Art*. Cambridge University Press.
- Genette, Gérard. 1992. *The Architext: An Introduction*. Quantum Books. University of California Press.
- Goetze, Trystan S. 2024. "AI Art Is Theft: Labour, Extraction, and Exploitation, Or, On the Dangers of Stochastic Pollocks." *arXiv.Org* (Ithaca), ahead of print. <https://doi.org/10.48550/arxiv.2401.06178>.
- Goldman, Alan H. 1995. *Aesthetic Value*. Focus Series. Westview Press.
- Goldsmith, Kenneth. 2011. *Uncreative Writing: Managing Language in the Digital Age*. Columbia University Press.
- Goodman, Nelson. 1976. *Languages of Art: An Approach to a Theory of Symbols*. 2d ed. Hackett.
- Gracia, Jorge. 1995. *A Theory of Textuality: The Logic and Epistemology*. SUNY Press.
- Gracia, Jorge. 1996. *Texts: Ontological Status, Identity, Author, Audience*. State University of New York Press.
- Hayles, Katherine. 2018. "Literary Texts as Cognitive Assemblages: The Case of Electronic Literature." *Electronic Book Review*, August 5.

<https://electronicbookreview.com/essay/literary-texts-as-cognitive-assemblages-the-case-of-electronic-literature/>.

Helliwell, Alice. 2023. "Art-ificial: The Philosophy of AI Art." Phd, University of Kent,.
<https://doi.org/10.22024/UniKent/01.02.105246>.

Helliwell, Alice. 2025. "AI as an Artistic Medium." *Art Style* 1 (16): 31–53.

Hessel, Jack, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. 2022.
"CLIPScore: A Reference-Free Evaluation Metric for Image Captioning."
arXiv:2104.08718. Preprint, arXiv, March 23. <http://arxiv.org/abs/2104.08718>.

Hirsch, E. D. 1967. *Validity in Interpretation*. Yale University Press.

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. 2020. "Denoising Diffusion Probabilistic Models."
arXiv:2006.11239. Preprint, arXiv, December 16.
<https://doi.org/10.48550/arXiv.2006.11239>.

Hong, Joo-Wha, and Nathaniel Ming Curran. 2019. "Artificial Intelligence, Artists, and Art: Attitudes Toward Artwork Produced by Humans vs. Artificial Intelligence." *ACM Trans. Multimedia Comput. Commun. Appl.* 15 (2s): 58:1-58:16.
<https://doi.org/10.1145/3326337>.

Irvin, Sherri. 2006. "Authors, Intentions and Literary Meaning." *Philosophy Compass* 1 (2): 114–28.

Irwin, William. 1999. *Intentionalist Interpretation: A Philosophical Explanation and Defense*. Greenwood Press.

- Irwin, William. 2015. "Authorial Declaration and Extreme Actual Intentionalism: Is Dumbledore Gay?" *Journal of Aesthetics and Art Criticism* 73 (2): 141–47.
- Ivanova, Milena. 2025. "AI, Art and Morality." *AI and Ethics* 5 (4): 4269–78.
<https://doi.org/10.1007/s43681-025-00735-3>.
- Johnston, David Jhave. 2019. *ReRites: Raw Output/Responses*. Anteism Books.
- Juhl, P. D. 1980. *Interpretation: An Essay in the Philosophy of Literary Criticism*. Princeton University Press.
- Kant, Immanuel. (1790) 2009. *The Critique of Judgement*. Edited by Nicholas Walker. Translated by James Creed Meredith. Oxford University Press.
- Kelly, Sean. 2019. "A Philosopher Argues That an AI Can't Be an Artist." MIT Technology Review. <https://www.technologyreview.com/2019/02/21/239489/a-philosopher-argues-that-an-ai-can-never-be-an-artist/>.
- Khosrowi, Donal, Finola Finn, and Elinor Clark. 2025. "Engaging the Many-Hands Problem of Generative-AI Outputs: A Framework for Attributing Credit." *AI and Ethics* 5 (5): 4495–513. <https://doi.org/10.1007/s43681-024-00440-7>.
- Lawson-Tancred, Jo. 2023. "In Pictures: Black Artists Use A.I. to Make Work That Reveals the Technology's Inbuilt Biases for a New Online Show." On View. *Artnet News*, July 3. <https://news.artnet.com/art-world/in-pictures-black-artists-use-a-i-to-make-work-that-reveals-the-technologys-inbuilt-biases-for-a-new-online-show-2322550>.

- Levinson, Jerrold. 1979. "Defining Art Historically." *The British Journal of Aesthetics* 19 (3): 232–50. <https://doi.org/10.1093/bjaesthetics/19.3.232>.
- Levinson, Jerrold. 1990. *Music, Art, and Metaphysics : Essays in Philosophical Aesthetics*. Cornell University Press.
- Levinson, Jerrold. 1996. "Intention and Interpretation in Literature." In *The Pleasures of Aesthetics: Philosophical Essays*. Cornell University Press.
- Li, Junnan, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. "BLIP-2: Bootstrapping Language-Image Pre-Training with Frozen Image Encoders and Large Language Models." arXiv:2301.12597. Preprint, arXiv, June 15. <http://arxiv.org/abs/2301.12597>.
- Li, Junnan, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. "BLIP: Bootstrapping Language-Image Pre-Training for Unified Vision-Language Understanding and Generation." arXiv:2201.12086. Preprint, arXiv, February 15. <http://arxiv.org/abs/2201.12086>.
- Li, Runjia, Shuyang Sun, Mohamed Elhoseiny, and Philip Torr. 2023. "OxfordTVG-HIC: Can Machine Make Humorous Captions from Images?" arXiv:2307.11636. Preprint, arXiv, July 21. <https://doi.org/10.48550/arXiv.2307.11636>.
- Lin, Szu-Yen. 2018. "Interpretation and the Implied Author: A Descriptive Project." *The Southern Journal of Philosophy* 56 (1): 83–100. <https://doi.org/10.1111/sjp.12269>.
- Lin, Szu-Yen. 2023. "Defending the Hypothetical Author." *The British Journal of Aesthetics* 63 (4): 579–99. <https://doi.org/10.1093/aesthj/ayac062>.

Longworth, Francis, and Andrea Scarantino. 2010. “The Disjunctive Theory of Art: The Cluster Account Reformulated.” *The British Journal of Aesthetics* 50 (2): 151–67.

<https://doi.org/10.1093/aesthj/ayq001>.

Lopes, Dominic. 2004. *Understanding Pictures*. Oxford University Press.

<https://doi.org/10.1093/acprof:oso/9780199272037.001.0001>.

Lopes, Dominic. 2010. *A Philosophy of Computer Art*. Routledge.

Lopes, Dominic. 2014. *Beyond Art*. Oxford University Press.

Lopes, Dominic. 2016. *Four Arts of Photography: An Essay in Philosophy*. Wiley.

Lopes, Dominic. 2018. *Being for Beauty: Aesthetic Agency and Value*. Oxford University Press.

MacBeth, George. 2023. “‘A Shifting Surface World’: The Techno-Graphomania of David Jhave Johnston’s *ReRites*.” *Electronic Book Review*, April 2.

<https://electronicbookreview.com/essay/a-shifting-surface-world-the-techno-graphomania-of-david-jhave-johnstons-rerites/>.

Mag Uidhir, Christy. 2011. “Minimal Authorship (of Sorts).” *Philosophical Studies* 154 (3): 373–87. <https://doi.org/10.1007/s11098-010-9525-0>.

Marcus, Gary, and Reid Southen. 2024. “Generative AI Has a Visual Plagiarism Problem.” *IEEE Spectrum*. <https://spectrum.ieee.org/midjourney-copyright>.

Mazzone, Marian, and Ahmed Elgammal. 2019. “Art, Creativity, and the Potential of Artificial Intelligence.” *Arts* 8 (1): 26. <https://doi.org/10.3390/arts8010026>.

- Meynell, Hugo. 1986. *The Nature of Aesthetic Value*. SUNY Press.
- Miller, Mara. 1993. *The Garden as an Art*. State University of New York Press.
- Moore, George. 1951. *Principia Ethica*. University Press.
- Moore, Joseph. 2025. “The Antinomy of Artificial Art.” *Philosophical Studies*, ahead of print, October 23. <https://doi.org/10.1007/s11098-025-02430-9>.
- Mordvintsev, Alexander, Christopher Berner, and Mike Tyka. 2015. *Inceptionism: Going Deeper into Neural Networks*. June 18. <https://research.google/blog/inceptionism-going-deeper-into-neural-networks/>.
- Nannicelli, Ted. 2025. “Mass AI-Art: A Moderately Skeptical Perspective.” *The Journal of Aesthetics and Art Criticism*, July 17, kpaf026. <https://doi.org/10.1093/jaac/kpaf026>.
- Narasimhaswamy, Supreeth, Uttaran Bhattacharya, Xiang Chen, Ishita Dasgupta, Saayan Mitra, and Minh Hoai. 2024. “HanDiffuser: Text-to-Image Generation With Realistic Hand Appearances.” arXiv:2403.01693. Preprint, arXiv, April 21. <https://doi.org/10.48550/arXiv.2403.01693>.
- Nehamas, Alexander. 1981. “The Postulated Author: Critical Monism as a Regulative Ideal.” *Critical Inquiry* (Chicago) 8 (1): 133–49. <https://doi.org/10.1086/448144>.
- Nehamas, Alexander. 1986. “What an Author Is.” *The Journal of Philosophy* 83 (11): 685–91.
- Nehamas, Alexander. 2007. *Only a Promise of Happiness: The Place of Beauty in a World of Art*. Princeton University Press.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. University Press.

Olsen, Stein Haugom. 1982. "The 'Meaning' of a Literary Work." *New Literary History* 14 (1): 13–32. <https://doi.org/10.2307/468955>.

Ouyang, Long, Jeff Wu, Xu Jiang, et al. 2022. "Training Language Models to Follow Instructions with Human Feedback." arXiv:2203.02155. Preprint, arXiv, March 4. <https://doi.org/10.48550/arXiv.2203.02155>.

Parrish, Allison. 2018. *Articulations*. Counterpath.

Parshall, Allison. 2023. "How This AI Image Won a Major Photography Competition." *Scientific American*, April 21. <https://www.scientificamerican.com/article/how-my-ai-image-won-a-major-photography-competition/>.

Paul, Elliot Samuel, and Dustin Stokes. 2018. "Attributing Creativity." In *Creativity and Philosophy*, edited by Berys Gaut and Matthew Kieran. Routledge.

Petsiuk, Vitali, and Kate Saenko. 2024. "Concept Arithmetics for Circumventing Concept Inhibition in Diffusion Models." arXiv:2404.13706. Preprint, arXiv, April 21. <https://doi.org/10.48550/arXiv.2404.13706>.

Propp, Vladimir. 1968. *Morphology of the Folktale*. Second Edition. Bibliographical and Special Series of the American Folklore Society ; v. 9. University of Texas Press.

- Radford, Alec, Jong Wook Kim, Chris Hallacy, et al. 2021. “Learning Transferable Visual Models From Natural Language Supervision.” arXiv:2103.00020. Preprint, arXiv. <http://arxiv.org/abs/2103.00020>.
- Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. “Improving Language Understanding by Generative Pre-Training.” <https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035>.
- Radford, Alec, Jeff Wu, R. Child, D. Luan, Dario Amodei, and I. Sutskever. 2019. “Language Models Are Unsupervised Multitask Learners.” <https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe>.
- Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. “Hierarchical Text-Conditional Image Generation with Clip Latents.” arXiv:2204.06125v1. Preprint, arXiv.
- Ramsden, Mel. 1975. “On Practice.” *The Fox*, no. 1.
- Richards, Ivor Armstrong. 1965. *The Philosophy of Rhetoric*. With Bryn Mawr College. The Mary Flexner Lectures, 1936. Oxford University Press.
- Richter, Hans. 1965. *Dada: Art and Anti-Art*. McGraw-Hill.

- Ridoy, Md Alif Rahman, M. Mahmud Hasan, and Shovon Bhowmick. 2024. "Compressed Image Captioning Using CNN-Based Encoder-Decoder Framework." arXiv:2404.18062. Preprint, arXiv, April 27. <https://doi.org/10.48550/arXiv.2404.18062>.
- Rini, Regina. 2024. "The Artificial Sublime." <https://philpapers.org/rec/RINTAS-2>.
- Ryle, Gilbert. (1949) 2002. *The Concept of Mind*. University of Chicago Press.
- Schmidt, João Vitor. 2025. "Can Artificial Agents Be Authors?" *Philosophy & Technology* 38 (1): 13. <https://doi.org/10.1007/s13347-025-00839-y>.
- Schuhmann, Christoph, Romain Beaumont, Richard Vencu, et al. 2022. "LAION-5B: An Open Large-Scale Dataset for Training next Generation Image-Text Models." arXiv:2210.08402. Preprint, arXiv, October 15. <https://doi.org/10.48550/arXiv.2210.08402>.
- Scruton, Roger. 1981. "Photography and Representation." *Critical Inquiry* 7 (3): 577–603.
- Searle, John. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417–24. <https://doi.org/10.1017/S0140525X00005756>.
- Sellors, C. Paul. 2007. "Collective Authorship in Film." *Journal of Aesthetics and Art Criticism* 65 (3): 263–71. <https://doi.org/10.1111/j.1540-594X.2007.00257.x>.
- Seltmann, Darren, and Robbie Chater. 2002. "The Avalanches." *Sound on Sound* (Cambridge, UK). <https://www.soundonsound.com/people/the-avalanches>.
- Shaffi, Sarah. 2023. "'It's the Opposite of Art': Why Illustrators Are Furious about AI." Art and Design. *The Guardian*, January 23.

<https://www.theguardian.com/artanddesign/2023/jan/23/its-the-opposite-of-art-why-illustrators-are-furious-about-ai>.

Singh, Jaskirat, and Liang Zheng. 2023. "Divide, Evaluate, and Refine: Evaluating and Improving Text-to-Image Alignment with Iterative VQA Feedback." arXiv:2307.04749. Preprint, arXiv, December 5. <http://arxiv.org/abs/2307.04749>.

Smith, Murray. 2006. "My Dinner with Noël; or, Can We Forget the Medium?" *Film Studies* (Manchester), no. 8: 140–48.

Smuts, Aaron. 2005. "Are Video Games Art?" *Contemporary Aesthetics* 3: 1–11.

Smuts, Aaron. 2009. "Art and Negative Affect." *Philosophy Compass* (Oxford, UK) 4 (1): 39–55. <https://doi.org/10.1111/j.1747-9991.2008.00199.x>.

Sohl-Dickstein, Jascha, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics." arXiv:1503.03585. Preprint, arXiv, November 18. <https://doi.org/10.48550/arXiv.1503.03585>.

Stecker, Robert. 2003. *Interpretation and Construction: Art, Speech, and the Law*. Blackwell Publishing.

Steiner, Wendy. 2001. *Venus in Exile: The Rejection of Beauty in Twentieth-Century Art*. University of Chicago Press.

Stillinger, Jack. 1991. *Multiple Authorship and the Myth of Solitary Genius*. Oxford University Press.

- Stinson, Catherine. 2022. "Algorithms Are Not Neutral." *AI and Ethics* 2 (4): 763–70.
<https://doi.org/10.1007/s43681-022-00136-w>.
- Stinson, Catherine. 2024. "What Are Large Language Models Made Of?" In *Leading the Way: Envisioning the Future of Higher Education*, edited by M. E. Norris and S. E. Smith.
- Stock, Kathleen. 2017. *Only Imagine: Fiction, Interpretation and Imagination*. Oxford University Press. <https://doi.org/10.1093/oso/9780198798347.001.0001>.
- Strohl, Matthew. 2012. "Horror and Hedonic Ambivalence." *The Journal of Aesthetics and Art Criticism* (Malden, USA) 70 (2): 203–12. <https://doi.org/10.1111/j.1540-6245.2012.01512.x>.
- Strohl, Matthew S. 2022. *Why It's OK to Love Bad Movies*. Why It's OK: The Ethics and Aesthetics of How We Live. Routledge, Taylor & Francis Group.
- Terrone, Enrico. 2018. "Appearance and History: The Autographic/Allographic Distinction Revisited." *The British Journal of Aesthetics* 58 (1): 71–87.
<https://doi.org/10.1093/aesthj/ayx034>.
- Theis, Lucas. 2024. "What Makes an Image Realistic?" arXiv:2403.04493. Preprint, arXiv, March 11. <https://doi.org/10.48550/arXiv.2403.04493>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, et al. 2017. "Attention Is All You Need." arXiv:1706.03762. Preprint, arXiv, December 5.
<https://doi.org/10.48550/arXiv.1706.03762>.

- Vlaad, Sofie. 2024. "A Portrait of the Artist as a Young Algorithm." *Ethics and Information Technology* 26 (3): 58. <https://doi.org/10.1007/s10676-024-09796-0>.
- Vlaad, Sofie. 2025. "Texts without Authors: Ascribing Literary Meaning in the Case of AI." *The Journal of Aesthetics and Art Criticism* 83 (1): 4–11. <https://doi.org/10.1093/jaac/kpae047>.
- Vlaad, Sofie, and Elliot Samuel Paul. Forthcoming. "Appreciating AI Art: Aesthetic Vices, Virtues, and Values." *Journal of Aesthetics and Art Criticism*.
- Walton, Kendall. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.
- Walton, Kendall. 2008. *Marvelous Images: On Values and the Arts*. Oxford University Press.
- Walton, Kendall L. 1970. "Categories of Art." *The Philosophical Review* (Boston) 79 (3): 334–67. <https://doi.org/10.2307/2183933>.
- Wang, Zhijie, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. "PromptCharm: Text-to-Image Generation through Multi-Modal Prompting and Refinement." *Proceedings of the CHI Conference on Human Factors in Computing Systems*, May 11, 1–21. <https://doi.org/10.1145/3613904.3642803>.
- Williams, Zoe. 2023. "'AI Isn't a Threat' – Boris Eldagsen, Whose Fake Photo Duped the Sony Judges, Hits Back." Art and Design. *The Guardian*, April 18. <https://www.theguardian.com/artanddesign/2023/apr/18/ai-threat-boris-eldagsen-fake-photo-duped-sony-judges-hits-back>.

- Wojtkiewicz, Kathryn. 2023. "How Do You Solve a Problem like DALL-E 2?" *The Journal of Aesthetics and Art Criticism* 81 (4): 454–67. <https://doi.org/10.1093/jaac/kpad046>.
- Yang, Ling, Zhilong Zhang, Yang Song, et al. 2023. "Diffusion Models: A Comprehensive Survey of Methods and Applications." arXiv:2209.00796v10. Preprint, arXiv.
- Young, Nick, and Enrico Terrone. 2025. "Growing the Image: Generative Ai and the Medium of Gardening." *Philosophical Quarterly*, ahead of print. <https://doi.org/10.1093/pq/pqae120>.
- Zangwill, Nick. 2013. "Clouds of Illusion in the Aesthetics of Nature." *The Philosophical Quarterly* (Oxford, UK and Boston, USA) 63 (252): 576–96.
<https://doi.org/10.1111/1467-9213.12050>.
- Zeimbekis, John. 2012. "Digital Pictures, Sampling, and Vagueness: The Ontology of Digital Pictures." *The Journal of Aesthetics and Art Criticism* 70 (1): 43–53.
<https://doi.org/10.1111/j.1540-6245.2011.01497.x>.
- Ziff, Paul. 1953. "The Task of Defining a Work of Art." *The Philosophical Review* 62 (1): 58–78.
- Ziff, Paul. 1984. *Antiaesthetics: An Appreciation of the Cow with the Subtile Nose*. D. Reidel Publishing Company.

Artworks and Images Cited

9/11 Gender Reveal. N.d. Synthetic image.

https://www.reddit.com/r/crayon/comments/vjytot/911_gender_reveal/. Licensed under the Creative Commons (CC0).

Allen, Jason. *Théâtre D'Opéra Spatial*. 2022. Synthetic image.

<https://web.archive.org/web/20251011160722/https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>. Licensed under the Creative Commons (CC0).

Atairu, Minnie. *Blonde Braids Study II*. 2023. Synthetic image.

<https://www.standard.co.uk/lifestyle/beauty/somerset-house-virtual-beauty-exhibition-b1216449.html>. Licensed under the Creative Commons (CC0).

Eldagsen, Boris. *The Electrician*. 2023. Synthetic image.

<https://www.theguardian.com/artanddesign/2023/apr/18/ai-threat-boris-eldagsen-fake-photo-duped-sony-judges-hits-back>. Licensed under the Creative Commons (CC0).

Mordvintsev, Alex. *Father Cat*. 2015. Synthetic image.

<https://foundation.app/mint/eth/0xD156EAb8d01F4165e5fa655CB87f17BE15e5292e/2>. Licensed under the Creative Commons (CC0).

National Park Services. *Six Grandfathers*. 1905. Photograph.

<https://www.nationalparkstraveler.org/2009/06/mount-who-how-did-famous-park-get-its-name>. Licensed under the Creative Commons (CC0).