NOTE

Copyright Infringement in AI-Generated Artworks

Jessica L. Gillotte[*]

TABLE OF CONTENTS

 J.D., 2020, University of California, Davis, School of Law; B.A., 2014, University of California, Berkeley. I am deeply grateful to Professors Elizabeth L. Rosenblatt, William S. Dodge, and Menesh S. Patel for their advice, guidance, and encouragement during the process of writing this Note. Thank you to Dillon Jackson for his thoughtful comments and fantastic editing, to the editors and members of the *UC Davis Law Review* for their tireless work, and to Sue Jones for her brilliance. Thank you also to my parents and sister for their unending support.

INTRODUCTION

Rembrandt van Rijn is generally considered one of the most important painters in history. He is famous for his use of light and shadow in his paintings and was prolific in producing portraits over a forty-year career.[1] Rembrandt completed his last painting in 1669, the year he died.[2] Over 300 years later, Dutch engineers created an artificial intelligence ("AI") program that "learned" the master's style through machine learning techniques and generated a portrait that resembled a Rembrandt original.[3] This project highlighted AI's remarkable capacity to generate expressive artworks.

Rembrandt's paintings are not currently protected by copyright,[4] so the Dutch engineers could freely digitize and use his art as input data to train the AI program to produce a similar-looking work. However, consider a situation where engineers train an AI program using copyrighted works. Does using copyrighted works to train an AI create infringement liability? Copyright law currently does not answer this question, but the growing use of AI in creating expressive artworks makes it interesting and important.[5]

---

[1] *See* Esther-Lee Marcus & A. Mark Clarfield, *Rembrandt's Late Self-Portrait: Psychological and Medical Aspects*, 55 INT'L J. AGING & HUMAN DEV. 25, 25 (2002).

[2] *See* Roderick Conway Morris, *Rembrandt's Turbulent Final Years*, N.Y. TIMES (Dec. 4, 2014), https://www.nytimes.com/2014/12/04/arts/international/rembrandts-turbulent-final-years.html [https://perma.cc/Z29T-KLS6]; *Self Portrait at the Age of 63, Rembrandt, 1669*, NAT'L GALLERY, https://www.nationalgallery.org.uk/paintings/rembrandt-self-portrait-at-the-age-of-63 (last visited Mar. 19, 2020) [https://perma.cc/A8NF-R5UL].

[3] *See* THE NEXT REMBRANDT, https://www.nextrembrandt.com (last visited Mar. 19, 2020) [https://perma.cc/7BG6-QJWH]. With respect to artificial intelligence, "learning" refers to an AI program's iterative process of identifying patterns and features in data and using insights to perform computing tasks such as finding meaning from unidentified data or identifying the content of images or audio. *See How Artificial Intelligence Works*, SAS, https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html#howitworks (last visited May 16, 2020) [https://perma.cc/G4VB-XM98]; *see also infra* notes 14–15 and accompanying text.

[4] *See* ANDREAS RAHMATIAN, COPYRIGHT AND CREATIVITY: THE MAKING OF PROPERTY RIGHTS IN CREATIVE WORKS 97-98 (2011).

[5] This Note seeks to answer this question with respect to AI-generated artworks for two reasons. First, focusing on one medium of expression makes the analysis more concrete. Second, there are more readily-available examples of AI-generated artwork in the public domain than AI-generated works of literature or music. *See, e.g.*, AIArtists.org, https://aiartists.org/ (last visited Mar. 19, 2020) [https://perma.cc/2BDS-B55J] (listing over twenty-five artists creating artworks using AI). Still, many AI-generated works of literature exist and continue to be produced, including fictional works and non-fictional research works. *See, e.g.*, Gregory Barber, *Text-Savvy AI Is Here to Write Fiction*, WIRED (Nov. 23, 2019, 7:00 AM), https://www.wired.com/story/nanogenmo-ai-novels-gpt2/ [https://perma.cc/A99R-KDP9]; Brian Merchant, *When an*

In *Twentieth Century Music Corp. v. Aiken*, the U.S. Supreme Court stated that "[w]hen technological change has rendered its literal terms ambiguous, the Copyright Act must be construed in light of [its] basic purpose." The basic purpose of the Copyright Act is set out in the U.S. Constitution's Intellectual Property Clause ("IP Clause"), which authorizes Congress "[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."[6] As the Supreme Court has explained in case law, the limited exclusive rights granted are meant to incentivize production and subsequent public dissemination of an author's works.[7] Using AI to create art comports with this constitutional purpose by increasing the volume of works produced, thereby making the works accessible to more people. But AI-generated artwork may also undermine copyright law's purpose of incentivizing *human* authors[8] to create new, original works, rather than

AI Goes Full Jack Kerouac, ATLANTIC (Oct. 1, 2018), https://www.theatlantic.com/technology/archive/2018/10/automated-on-the-road/571345/ [https://perma.cc/ZF28-9UV5]; James Vincent, *The First AI-Generated Textbook Shows What Robot Writers Are Actually Good At*, VERGE (Apr. 10, 2019, 11:55 AM), https://www.theverge.com/2019/4/10/18304558/ai-writing-academic-research-book-springer-nature-artificial-intelligence [https://perma.cc/VK2R-6CDC].

[6] *See* U.S. CONST. art. I, § 8, cl. 8.

[7] This conception of the justification of copyrights is embodied by the utilitarian theory, which is the theory dominant in the United States. The utilitarian theory holds that copyright is justified because it incentivizes individuals to create new works of art, which, without copyright protection, would be easily and cheaply copied, thus depriving the original creator of the economic benefits of their work. *See* JEANNE C. FROMER & CHRISTOPHER JON SPRIGMAN, COPYRIGHT LAW: CASES AND MATERIALS v. 1.0, at 9-10 (2019); Shyamkrishna Balganesh, *Foreseeability and Copyright Incentives*, 122 HARV. L. REV. 1569, 1576-77 (2009); Jeanne C. Fromer, *An Information Theory Of Copyright Law*, 64 EMORY L.J. 71, 74 (2014) (citing Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 558 (1985) and 122 Cong. Rec. 2834-35 (1976) (statement of Sen. John McClellan)); William M. Landes & Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 J. LEGAL STUD. 325, 326 (1989); *see also infra* notes 64–66 and accompanying text.

[8] The copyright statute states that authors include creators of any kind of copyrightable work. *See* 17 U.S.C. § 102 (2018) (enumerating categories of copyrightable works of authorship). While the statute does not state whether a computer program can be an author, at least one court, the Ninth Circuit, has held that animals may not be authors because they do not have Article III standing to sue for copyright infringement. *See infra* note 61 and accompanying text. Further, the U.S. Copyright Office has stated that only human beings can be authors. U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES §§ 306, 313.2 (3d ed. 2017) [hereinafter COMPENDIUM] ("To qualify as a work of 'authorship' a work must be created by a human being." (citing Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 58 (1884))). The U.S. Copyright Office has also stated that it will not register works "produced by a machine or mere mechanical process that operates randomly or

merely incentivizing the *creation* of new, original works.[9] When AI-generated works directly compete with those of human authors, the latter may eventually stop creating as they see the market for their output shrink.

This Note examines potential copyright infringement issues arising from AI-generated artwork and argues that, under current copyright law, an engineer may use copyrighted works to train an AI program to generate artwork without incurring infringement liability.

The Note proceeds in three parts. Part I provides an overview of how AI is used to generate artwork and provides background on the issues of copyrightability and ownership of AI-generated works. Part II turns to issues of infringement. It first dispenses with the argument that the use of copyrighted works in machine learning is considered a non-infringing "operational" use. The Part then describes a circuit split as to what infringing copies are and presents two scenarios that could create infringing copies: (1) using unauthorized reproductions to form a training set, and (2) creating unauthorized intermediate copies during the process of machine learning.

Finally, Part III argues that even if infringement occurs during machine learning, training AI with copyrighted works would likely be excused by the fair use doctrine. The Note thus concludes that despite some calls for the Copyright Act to be amended,[10] there may be no need to address expressly by statute the copyright infringement issues that arise from the use of AI to generate artworks.

I. INTERACTIONS BETWEEN ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW

Artificial intelligence raises a number of interesting questions of copyright law. This Part provides background for the consideration of copyright infringement issues that are the focus of this Note. Part I.A provides an overview of AI, machine learning, and neural networks, explaining how such technologies might use copyrighted materials. Part

automatically *without any creative input or intervention* from a human author." *Id.* § 313.2 (emphasis added). Note, however, that registration with the U.S. Copyright Office is not necessary to obtain a copyright in a work. *Id.* § 202 ("[R]egistration is not required for a work to be protected by copyright . . . .").

[9] *See* Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975) ("The immediate effect of our copyright law is to secure a fair return for an 'author's' creative labor. But the ultimate aim is, by this incentive, to stimulate artistic creativity for the general public good.").

[10] *See, e.g.*, CHEN ET AL., SAMUELSON L. TECH. & PUBLIC POL'Y CLINIC, PROVIDING AN INCIDENTAL COPIES EXEMPTION FOR SERVICE PROVIDERS AND END-USERS 6, 8, 10-16 (2011).

I.B then briefly considers copyright issues other than infringement, such as the copyrightability of AI-generated works and who might own the copyrights in such works.

A. Overview of AI, Machine Learning, and Artificial Neural Networks

Artificial intelligence is a field of computer science involved with developing a computer's capacity to behave as an intelligent entity.[11] Intelligent entities can perform both technical functions, such as finding an optimal math solution, and functions traditionally associated only with the human brain, such as processing natural language.[12] There are many sub-fields of AI research and this Note focuses on machine learning. In broad terms, machine learning is a process by which AI extrapolates patterns from large quantities of data and uses those patterns to learn the constraints of the output it is expected to produce without being explicitly programmed to produce it.[13] During machine learning, the AI program receives feedback and refines its underlying algorithm to improve its performance of the defined task over time.[14]

An AI program can learn by receiving feedback from two alternative methods of training: supervised learning and unsupervised learning. In supervised learning, the AI is given labeled training data — a set of input-output pairs — such as a set of images of a flower, with each image labeled *flower*.[15] The AI then constructs an algorithm that accurately maps the input images to the output label.[16] Since the AI program is in possession of a previously labeled target output value (here, *flower*), the program immediately adjusts its algorithm to produce an output that is as close as possible to the correct output label.[17] In unsupervised learning, the AI program observes patterns in the input data without explicit feedback or labels and instead refines its algorithms iteratively by comparing its performances over time.[18]

---

[11] STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 1 (3d ed. 2010).

[12] *See id.* at 1-2.

[13] *See* M. Tim Jones, *A Beginner's Guide to Artificial Intelligence, Machine Learning, and Cognitive Computing*, IBM DEVELOPERWORKS (Oct. 12, 2018), https://developer.ibm.com/articles/cc-beginner-guide-machine-learning-ai-cognitive/#machine-learning [https://perma.cc/TTQ8-UHN4]. Machine learning has roots in statistics and mathematical optimization, wherein computers "learn" in order to predict, analyze, and mine data. *Id.*

[14] *See* RUSSELL & NORVIG, *supra* note 11, at 693.

[15] *See id.* at 695.

[16] *See id.*

[17] *See id.*

[18] *See id.* at 694-95.

One approach to machine learning involves using artificial neural networks. A neural network is a structured arrangement of layers of artificial neurons — mathematical units that contain simple mathematical functions.[19] The artificial neurons mimic biological neurons by transmitting a signal to the next artificial neuron when a linear combination of the input it receives exceeds a certain threshold value.[20] Neural networks range in complexity. A simple neural network can have only one layer and perform simple classification tasks.[21] More complex neural networks contain many layers of neurons and, in addition to classification tasks, can perform regression tasks such as generating a new image from a corpus of input data images.[22] The convolutional neural network ("CNN"), created by Yann LeCun in 1995, is an example of a complex neural network.[23] A CNN is a multi-layer neural network that makes use of backpropagation, which involves successive adjustments to the AI's algorithm based on the closeness of the last iterative output to the ideal.[24] In other words, the AI measures the error in its last performance to know how to improve its next performance.

[19] *See id.* at 728.

[20] *See id.* at 727-28.

[21] An example of a simple neural network is the perceptron, a supervised learning algorithm wherein all the inputs are directly connected to the outputs. The perceptron receives input data and classifies the input as belonging to some category. It then adjusts its algorithm based on feedback based on the labeled training set. *See id.* at 728-30; Jones, *supra* note 13.

[22] Pankaj Mehta et al., *A High-Bias, Low-Variance Introduction to Machine Learning for Physicists*, 810 PHYS. REP. 1, 4 (2019).

[23] Yann LeCun, *LeNet-5, Convolutional Neural Networks*, http://yann.lecun.com/exdb/lenet/ (last visited Jan. 30, 2020) [https://perma.cc/B6UC-UVUM]; *see also* Jones, *supra* note 13.

[24] In more detailed terms, the process of backpropagation in machine learning involves many steps. In the first step, the AI is given numerical input data which comprises the first layer of artificial neurons. The input data is multiplied by a random weight and then added to a random number called the *bias*. If the sum of the weight and bias exceeds a defined threshold, a neuron in the second layer is "fired," and contains the value of the calculated sum. This second-layer neuron is also multiplied by some random *weight* and added to the random *bias*, causing a neuron in the third layer to "fire" if it exceeds some defined value. The consecutive firing of sequential layers of neurons eventually reaches the final layer, which presents the algorithm's particular output. *See generally* MICHAEL NIELSEN, NEURAL NETWORKS AND DEEP LEARNING, Ch. 2: How the Backpropagation Algorithm Works, http://neuralnetworksanddeeplearning.com/chap2.html (last visited Jan. 30, 2020) [https://perma.cc/ML8B-SVNF] (explaining the origination of the backpropagation algorithm and outlining the process); 3Blue1Brown, *What Is Backpropagation Really Doing? | Deep Learning, Chapter 3*, YOUTUBE (Nov. 3, 2017), https://www.youtube.com/watch?v=Ilg3gGewQ5U [https://perma.cc/6Q78-VFE5] (using visual concepts to explain the backpropagation process).

A more recent type of complex neural network is the Generative Adversarial Network ("GAN"), created by Ian J. Goodfellow in 2014.[25] GANs take a game theoretical approach to machine learning by making use of two simultaneously trained networks that are tasked with outperforming each other.[26] The first network, the generative model, begins with a sample of random data and generates a random output image.[27] Because the data used to generate the image is random, the first several images created by the GAN's generative model will appear crude and shapeless.[28] Then, the second network, the discriminative model, tries to determine whether the generative model's output image is generated or real.[29] Both networks are trained via backpropagation, and, as the generative model and discriminative model try to outmaneuver one another, the overall performance of the GAN improves.[30] Thus, over time, the GAN's generative model creates images that are more difficult to distinguish from the real ones, and the discriminative model becomes better at determining whether an image is real or generated. Eventually, the generated output images are no longer distinguishable from the real images.[31]

Unlike older generative neural networks that require intermediate copies to learn,[32] the GAN's generative model can create images without intermediate copies.[33] As Goodfellow writes, the GAN's generative model is "not being updated directly with data examples, but only with

[25] Ian J. Goodfellow et al., *Generative Adversarial Nets*, 27 PROC. NEURAL INFO. PROCESSING SYSTEMS 2672, 2672 (2014).

[26] *Id.*

[27] *See id.* (describing that the generative model captures a random data distribution).

[28] *See id.* (describing that the generative model aims to maximize the probability that the discriminative model mistakes a sample of the generative model as a sample of the training data distribution rather than the random data distribution).

[29] *Id.* at 2672 ("The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency.").

[30] *See id.* at 2673.

[31] *See id.* at 2672. Goodfellow et al. describes that a "unique solution" to this adversarial network exists where the generative model "recover[s] the training data distribution" and the discriminative model is "equal to ½ everywhere," meaning that the model is not able to tell whether the generative model is real or fake. *Id.*

[32] *See* Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 62, 62 n.97 (2017) (describing that during machine learning, deep auto-encoders "repeatedly encode and decode training data or elements thereof" (quoting KEVIN P. MURPHY, MACHINE LEARNING: A PROBABILISTIC PERSPECTIVE 1003-05 (2007))).

[33] *See* Goodfellow et al., *supra* note 25, at 2678.

[error calculations] flowing through the discriminator. . . . [The] components of the input are not copied directly into the generator's parameters."[34] Rather, the GAN's generative model creates the first image from random data and learns only from the discriminative model's feedback.[35]

Older generative neural networks such as the CNN require tens of thousands or perhaps even millions of images to train the AI program, while the GAN's discriminative model requires as few as 100-300 real images to learn how to distinguish real from generated images.[36] GANs therefore represent a forward leap in AI research, permitting the creation of more output with less input.[37] Nonetheless, while the GAN's generative model learns to create images without any *direct* input data, the discriminative model still requires input data.[38] Even with these advanced AI capabilities allowing computers to generate images without direct human involvement, computers still must learn from *something*, and the images they learn from may be copyrighted.

Artificial intelligence has been used to generate art for almost fifty years,[39] and huge strides in AI research over the past decade have enabled AI to create very complex artworks. Because of GANs' flexibility and effectiveness, this learning system is frequently used to create AI-generated artworks.[40] Such works have been published and posted

---

[34] *Id.*

[35] *See id.*

[36] *See* Martin Giles, *The GANfather: The Man Who's Given Machines the Gift of Imagination*, MIT TECH. REV. (Feb. 21, 2018), https://www.technologyreview.com/s/610253/the-ganfather-the-man-whos-given-machines-the-gift-of-imagination/ [https://perma.cc/6TE8-JMJ4].

[37] *See id.*

[38] *See* Goodfellow et al., *supra* note 25, at 2678.

[39] One of the earliest uses of AI to create art was a computer system created by Harold Cohen in 1973 called AARON. Chris Garcia, *Harold Cohen and AARON — A 40-Year Collaboration*, COMPUTER HISTORY MUSEUM (Aug. 23, 2016), http://www.computerhistory.org/atchm/harold-cohen-and-aaron-a-40-year-collaboration [https://perma.cc/FM7P-WDYB]. Cohen taught AARON to draw the way an adult might teach a child to draw, for example, by teaching a child to enclose line scribbles in a closed form. *See* Aaron Cohen, *How to Draw Three People in a Botanical Garden* (1988), https://pdfs.semanticscholar.org/d2dc/74733ecea185052135b336014d84a08a589c.pdf [https://perma.cc/AF9V-DVSS]. *See generally* Arthur R. Miller, *Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU?*, 106 HARV. L. REV. 977, 1047 (1993).

[40] *See* Kelsey Campbell-Dollaghan, *This Nude Portrait Was Generated by Algorithms*, FAST CO. (Oct. 1, 2018), https://www.fastcompany.com/90243942/this-award-winning-nude-portrait-was-generated-by-an-algorithm [https://perma.cc/9N9A-WLDZ] (describing artist Mario Klingemann's work in using more than one GAN to generate an image of a nude body and then to refine, or perform "transhancement," on the generated image).

online,[41] and displayed in physical spaces and galleries around the world.[42] As described in the Introduction, in 2016 a group of Dutch engineers created an AI program that generated a portrait entitled *The Next Rembrandt* that resembled a real Rembrandt portrait.[43] Through this process of machine learning, the AI generated a new portrait that many would mistake for a real Rembrandt painting.[44]

[41] *See, e.g.*, Mario Klingemann, QUASIMODO, http://quasimondo.com/ (last visited Mar. 19, 2020) [https://perma.cc/C99E-S85F]; Anna Ridler, *GANs in Art*, ANNA RIDLER, http://annaridler.com/gans-in-art/ (last visited Mar. 19, 2020) [https://perma.cc/SX3S-VN8N]; Helena Sarin (@glagolista), TWITTER, https://twitter.com/glagolista?lang=en (last visited Mar. 19, 2020) [https://perma.cc/ZVM9-U825] (showcasing several GAN-generated works that the author has produced).

[42] *See* Ridler, *supra* note 41 (explaining that art generated by GANs continue to appear in the "international fine arts scene," including at Ars Electronica 2017 and Serpentine Gallery Miracle Marathon 2017); *see, e.g.*, *Gradient Descent*, NATURE MORTE, http://naturemorte.com/exhibitions/gradientdescent/ (last visited Mar. 19, 2020) [https://perma.cc/E79G-4NMM] (describing an exhibition of art that has been created only using AI, featuring pioneers in AI-generated art such as Harshit Agrawal, Memo Akten, Jake Elwes, Mario Klingemann, Anna Ridler, Nao Tokui, and Tom White).

[43] *See The Next Rembrandt: Blurring the Lines Between Art, Technology and Emotion*, MICROSOFT NEWS CTR. EUR. (Apr. 13, 2016), https://news.microsoft.com/europe/features/next-rembrandt/ [https://perma.cc/EJ8P-4LA6]. The engineers first chose a subset of portraits deemed most representative of Rembrandt's oeuvre. *See Gathering the Data*, NEXT REMBRANDT, https://www.nextrembrandt.com/ (last visited Jan. 30, 2020) [https://perma.cc/7BG6-QJWH]. They then digitized the 346 chosen Rembrandt portraits with high resolution photography. *Id.* The subjects of this subset of portraits were white males, between thirty to forty years old, bearded, wearing black clothing with a white collar and a black hat, and facing to the right. *See Determining the Subject*, NEXT REMBRANDT, https://www.nextrembrandt.com/ (last visited Jan. 30, 2020) [https://perma.cc/7BG6-QJWH]. Then, the engineers used AI algorithms first to maximize the resolution and quality of the images and then to classify the most typical geometric patterns used in the Rembrandt portraits. *See Generating the Features*, NEXT REMBRANDT, https://www.nextrembrandt.com/ (last visited Jan. 30, 2020) [https://perma.cc/7BG6-QJWH]; *see also* Mark Brown, *'New Rembrandt' to be Unveiled in Amsterdam*, GUARDIAN (Apr. 5, 2016, 4:00 PM), https://www.theguardian.com/artanddesign/2016/apr/05/new-rembrandt-to-be-unveiled-in-amsterdam [https://perma.cc/6VZ6-GNDZ]. Finally, two algorithms determined "texture patterns on canvas surfaces and layers of paint" and the engineers used a multi-layer printing technique so that the portrait had visibly raised paint surfaces that mimicked Rembrandt's style. *See Bringing It to Life*, NEXT REMBRANDT, https://www.nextrembrandt.com/ (last visited Jan. 30, 2020) [https://perma.cc/7BG6-QJWH].

[44] *See A 'New' Rembrandt: From the Frontiers of AI and Not the Artist's Atelier*, NPR (Apr. 6, 2016, 10:23 PM), https://www.npr.org/sections/alltechconsidered/2016/04/06/473265273/a-new-rembrandt-from-the-frontiers-of-ai-and-not-the-artists-atelier [https://perma.cc/4LPG-696X] (noting that the project's head designer, Bas Korsten, did not himself think experts would believe *The Next Rembrandt* was an original Rembrandt painting).

More recently, a French art collective called Obvious generated its own painting with a GAN model that used open-source code written by AI artist Robbie Barrat.[45] The engineers inputted images of classical portraits into the GAN.[46] The generative and discriminative models then tried to outmaneuver each other until a painting, later entitled *Portrait of Edmond Belamy*, was produced.[47] In contrast to *The Next Rembrandt*, which was never offered for sale,[48] *Portrait of Edmond Belamy* commanded $432,500 at a Christie's auction in New York City on October 25, 2018.[49] The large sum of money has quickly focused the general public's attention on AI-generated art.[50]

With the increasing popularity of using AI to generate art, more copyrighted images may be used as training data. Those utilizing copyrighted input data to create AI-generated art may therefore encounter copyright infringement issues.

B. Aspects of Copyright Law Relevant to AI-Generated Works

Before turning to issues of infringement, it is helpful to understand how other aspects of copyright law apply to AI-generated artworks. The first Subsection explains what is required for copyrightability. Assuming that AI-generated works are copyrightable, the second

---

[45] *See* Obvious, *Obvious, Explained*, MEDIUM (Feb. 14, 2018), https://medium.com/@hello.obvious/ai-the-rise-of-a-new-art-movement-f6efe0a51f2e [https://perma.cc/G93Q-NXH5]. For the open-source code utilized by Obvious, see Robbie Barrat, *art-DCGAN*, GITHUB, https://github.com/robbiebarrat/art-DCGAN (last visited Jan. 26, 2020) [https://perma.cc/8K2T-7E3P].

[46] *See* Obvious, *supra* note 45.

[47] The Obvious engineers describe their process as follows. First, the engineers selected a large quantity of input images of "classical portraits" containing common visual features. Next, a generative model was trained to generate images containing the common visual feature. The discriminative model was trained to distinguish the generated images from the real images. The AI performed this process many times, refining the generative and discriminative models so that one may successfully prevail over the other. *See* Obvious, *supra* note 45.

[48] *The Next Rembrandt*, which was commissioned by ING Bank and completed in partnership with Microsoft, was displayed at an undisclosed space in Jordaan, Amsterdam. *See The Next Rembrandt: Blurring the Lines Between Art, Technology and Emotion*, *supra* note 43.

[49] *See* Gabe Cohn, *AI Art at Christie's Sells for $432,500*, N.Y. TIMES (Oct. 25, 2018), https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html [https://perma.cc/7JMP-ZSN2]; *Is Artificial Intelligence Set to Become Art's Next Medium?*, CHRISTIE'S (Dec. 12, 2018), https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx [https://perma.cc/QMC7-Z7V7].

[50] Noted that this sale was not without controversy. *See infra* note 206.

Subsection explores who might own the copyright in the AI-generated work.

1. Copyrightable Subject Matter

U.S. copyright law does not expressly address AI-generated works.[51] Nonetheless, statutory requirements of copyrightable subject matter, codified at 17 U.S.C. § 102, may be applied to a copyright analysis of AI-generated work. To be copyrightable, a work must be (1) an original work of authorship, and (2) fixed in any tangible medium of expression.[52] An original work of authorship must be independently created by the author and embody some minimal amount of creativity.[53] The work can require the use of a machine or other devices to view its expression — for example, films are copyrightable expression even when they are not viewable without the aid of a projector, television, computer, or other device.[54]

When the requirements of copyrightable subject matter are met, the author of the work is endowed with a bundle of independent exclusive rights for a limited time, including the rights to make reproductions; sell, rent, or give the work away; prepare new, derivative works; publicly perform the work; and publicly display the work.[55] These rights protect authors against copyists who seek to profit from an author's work without expending the resources, time, and energy that the author did in creating her work.

2. Ownership of the Copyright

Assuming that an AI-generated work is copyrightable, we turn to the question of who owns that copyright. The question is decades old. In 1974, the U.S. Congress established the Commission on New Technological Uses of Copyrighted Works ("CONTU") to study how computers fit into the U.S. copyright regime.[56] CONTU's study led to

[51] In contrast, for example, the United Kingdom has expressly addressed AI-generated works in its copyright laws. Copyright, Designs and Patents Act 1988, c. 48, § 9 (Eng.) (codifying that the person responsible for making the necessary arrangements to create the computer-generated work will be its owner).

[52] 17 U.S.C. § 102 (2019).

[53] *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 362-63 (1991) (holding that some minimal amount of creativity must be found to constitute original expression in copyrightable works).

[54] *See id.*

[55] 17 U.S.C. § 106 (2019).

[56] *See* Miller, *supra* note 39, at 979.

the enactment of the Copyright Act of 1976 ("Copyright Act"), the first update to the original Copyright Act of 1909.[57] However, left unresolved by CONTU and unaddressed in the Copyright Act was the question of copyright ownership in independent, computer-generated works. Such technology was not fathomable at the time,[58] but today computers can create expressive works that meet the elements of copyrightable subject matter,[59] and many scholars have analyzed this ownership question.

Both the U.S. Copyright Office and scholars agree that computer programs may not own copyrights because software has no legal personhood and therefore cannot own property.[60] This leaves four other candidates for ownership in computer-generated works: (1) the computer programmer, (2) the computer user, (3) both the programmer and the user as joint authors, or (4) simply no one.

The primary argument for the programmer's ownership is based on an augmented "works made for hire" doctrine.[61] The works made for hire doctrine allows an employer to own the copyright to a work created by her employee within the scope of the employee's employment.[62] However, the Supreme Court has interpreted "employment" as "the conventional relation of employer and employe[sic]."[63] Therefore, resolving the ownership issue using the works made for hire doctrine

---

[57] *See id.* at 979-80.

[58] *See id.* at 1046 (explaining that perhaps CONTU did not address this ownership question because even at the time of Professor Miller's writing, "the day when a computer is capable of creating *Ulysses* [was] not . . . approaching").

[59] *See* Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185, 1198-99 (1986) (stating that it is established in copyright law that "'original' means only that a work 'owes its origin' to any 'author,'" and that it does not mean the work has to be "startling, novel, or unusual" (quoting Alfred Bell & Co. v. Catalda Fine Arts, Inc., 191 F.2d 99, 102-03 (2d Cir. 1951))).

[60] *See* Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5, 21; U.S. COPYRIGHT OFFICE, COMPENDIUM, *supra* note 8, at § 306. A corollary to this position was illustrated in the so-called monkey selfie case, *Naruto v. Slater*, 888 F.3d 418 (9th Cir. 2018), wherein plaintiff PETA claimed that a photographer's publishing of selfies taken by a Macaque monkey constituted copyright infringement. There, the Ninth Circuit held that the monkey lacked statutory standing to sue under copyright law because "[t]he Copyright Act does not expressly authorize animals to file copyright infringement suits under the statute." *Id.* at 426.

[61] *See* Bridy, *supra* note 60, at 26-28.

[62] *See* 17 U.S.C. §§ 101, 201(b) (2019) (describing that works made for hire are works "prepared . . . within the scope of . . . employment" or is "a work specially ordered or commissioned for use as a contribution to a collective work").

[63] *Cmty. for Creative Non-Violence*, 490 U.S. at 740 (citing Kelley v. S. Pac. Co., 419 U.S. 318, 323 (1974)); *see also* Bridy, *supra* note 60, at 27 (citing Cmty. for Creative Non-Violence v. Reid, 490 U.S. 730, 731 (1989) (holding that "employee" under 17 U.S.C. § 101 is to be defined with reference to agency law)).

would require expanding "employment" to include the relationship between a programmer and her computer.

The argument that the computer user should own the copyright is based on the utilitarian theory of copyright law.[64] The utilitarian theory, which is the prevailing theory of copyright law in the United States,[65] finds support in U.S. copyright law and the Constitution's IP Clause and has been affirmed in many Supreme Court decisions.[66] It holds that the purpose of copyright law is to incentivize authors to create expressive works for the public benefit.[67] In line with this utilitarian purpose, granting the copyright to the computer user does incentivize the entity best-positioned to create and introduce AI-generated works to the public, although indirectly through the computer's programmer.[68]

The argument for why both the programmer and user should share the ownership of the copyright as joint authors is responsive to the fact that while both entities were necessary to create the AI-generated work,

[64] *See* Samuelson, *supra* note 59, at 1227 (arguing that recognizing the user as the owner of computer-generated works is consistent with the constitutional purposes of copyright law because "it has more potential to advance the pace of innovation than would be the case if no one was granted rights" (citing *Copyright and Technological Change: Hearings Before the Subcomm. on Courts, Civil Liberties & the Admin. of Justice of the H. Comm. on the Judiciary*, 98th Cong. (1983) at 8)).

[65] Legal regimes outside of the United States have adopted other theories. For example, the natural rights theory attributes value to the labor authors expend in producing works. *See* Lior Zemer, *The Making of a New Copyright Lockean*, 29 HARV. J.L. & PUB. POL'Y 891, 909 n.87 (2006) (stating that under this theory, human authors deserve rewards for the labor they put into their creations). Finally, the moral rights theory reasons that a work is inherently imbued with some aspect of the author's personality and must be protected from harms caused by theft or infringement. The moral rights theory, also called the personhood theory, has roots in French and German copyright law. *See* William Fisher, *Theories of Intellectual Property*, *in* NEW ESSAYS IN THE LEGAL AND POLITICAL THEORY OF PROPERTY 168, 174 (Stephen R. Munzer ed., 2001).

[66] *See* U.S. CONST. art. I, § 8, cl. 8 (Congress has the power "[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."); *see, e.g.*, N.Y. Times Co. v. Tasini, 533 U.S. 483, 520 (2001) (Stevens, J., dissenting) ("[C]opyright law demands that 'private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts.'" (quoting Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975))); Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 575 (1994) (explaining that the overriding purpose of copyright is "to promote the Progress of Science and useful Arts" (quoting U.S. CONST. art. I, § 8, cl. 8)); *see also* Margot Kaminski, *Authorship, Disrupted: AI Authors in Copyright and First Amendment Law*, 51 UC DAVIS L. REV. 589, 599 (2017).

[67] *See* sources cited *supra* note 7 and accompanying text.

[68] *See* Robert C. Denicola, *Ex Machina: Copyright Protection for Computer-Generated Works*, 69 RUTGERS U. L. REV. 251, 283-84 (2016) (explaining that users as authors is preferable both on practical and policy grounds).

neither could have created it on their own.[69] Finally, the argument for why no one should own the copyright is also based on the utilitarian theory of copyright law, as ownerless works can be freely and widely shared with the public.[70]

Resolving the uncertainty about ownership and the legal status of AI-generated works is certainly of interest to buyers and sellers of the works, but it is also of interest to current owners of copyrighted works that might be used as input data. Infringement questions arise when copyrighted works are used to train AI programs to generate art. The next Part analyzes such potential infringement scenarios and how courts might resolve them.

II. THE LEGAL STATUS OF AI-GENERATED WORKS REMAINS UNCERTAIN

This Note has so far considered the copyrightability and ownership of AI-generated works.[71] This Part focuses on the issue of copyright infringement, another area of legal uncertainty for AI-generated works.

When a copyright owner's independent exclusive rights are violated, the owner can sue for infringement under 17 U.S.C. § 501.[72] The copyright owner bears the burden of establishing a prima facie case for infringement.[73] To succeed on an infringement claim, the plaintiff must prove that (1) she owns of a valid copyright to her work, (2) the defendant engaged in unauthorized copying of original elements of the work, and (3) the copying was substantial and amounted to "an improper or unlawful appropriation"[74] that violates a statutory

[69] *See* Bruce E. Boyden, *Emergent Works*, 39 COLUM. J.L. & ARTS 377, 384-88 (2016) (explaining that delineating where a programmer's expression ends and the user's expression begins is a complicated task).

[70] *See, e.g.*, Ralph D. Clifford, *Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand Up?* , 71 TUL. L. REV. 1675, 1695 (1997).

[71] *See supra* Part I.B.

[72] *See* Sony Corp. v. Universal City Studios, Inc., 464 U.S. 417, 432-33 (1984).

[73] *See* Keeler Brass Co. v. Cont'l Brass Co., 862 F.2d 1063, 1066 (4th Cir. 1988).

[74] *See* Castle Rock Entm't, Inc. v. Carol Publ'g Grp., Inc., 150 F.3d 132, 137 (2d Cir. 1998) (citing Laureyssens v. Idea Grp., Inc., 964 F.2d 131, 139-40 (2d Cir. 1992)).

exclusive right.[75] Intent plays no part in the analysis.[76] Therefore, even if a defendant believes in good faith that she is not infringing, she may still be found liable if the elements of infringement are satisfied.[77] Conversely, defendants with an intent to infringe may avoid liability if those elements are not established.[78]

The first element of infringement, ownership of a valid copyright in the work, may be established by the plaintiff's proof of a registered copyright.[79] The second element, copying of original elements of the plaintiff's work, requires the plaintiff to prove that copying occurred and that the elements copied were expressions rather than ideas.[80] The third element, substantial similarity between the works, requires the plaintiff to show that an ordinary person would think the allegedly copied elements were substantially similar to the elements in the

[75] *See* 17 U.S.C. § 102 (2019) (setting out categories of copyrightable subject matter); *id.* § 411(a) (requiring as a prerequisite for bringing copyright infringement action the pre-registration or registration of a valid copyright); *Feist Publ'ns, Inc.*, 499 U.S. at 361 (setting out the elements of a copyright infringement claim); Tanksley v. Daniels, 902 F.3d 165, 172 (3d Cir. 2018) (explaining that especially where the allegedly infringing work is not highly technical in nature, the substantial similarity analysis is based on a layman's perspective); Peter Pan Fabrics, Inc. v. Martin Weiner Corp., 274 F.2d 487, 489 (2d Cir. 1960) (explaining that substantial similarity is established when an "ordinary observer, unless he set out to detect the disparities, would be disposed to overlook them, and regard their aesthetic appeal as the same").

[76] *See* Pinkham v. Sara Lee Corp., 983 F.2d 824, 829 (8th Cir. 1992). *But see* CoStar Grp., Inc. v. LoopNet, Inc., 373 F.3d 544, 548 (4th Cir. 2004) (explaining that in some instances, where the alleged infringer serves only as a passive conduit, "there should still be some element of volition or causation" required (quoting Religious Tech. Ctr. v. Netcom On-Line Commc'n Servs., Inc., 907 F. Supp. 1361, 1370 (N.D. Cal. 1995))).

[77] *See* Pye v. Mitchell, 574 F.2d 476, 481 (9th Cir. 1978) ("Indeed, even where the defendant believes in good faith that he is not infringing a copyright, he may be found liable."); Bright Tunes Music Corp. v. Harrisongs Music, Ltd., 420 F. Supp. 177, 181 (S.D.N.Y. 1976) (finding that the defendant infringed upon the copyright "even though subconsciously accomplished").

[78] *See* Lottie Joplin Thomas Tr. v. Crown Publishers, Inc., 456 F. Supp. 531, 537 (S.D.N.Y. 1977) ("An actual intent to infringe (or lack thereof) is immaterial to liability."); Pickwick Music Corp. v. Record Prods., Inc., 292 F. Supp. 39, 41 (S.D.N.Y. 1968); Screen Gems-Columbia Music, Inc. v. Mark-Fi Records, Inc., 256 F. Supp. 399, 402 (S.D.N.Y. 1966).

[79] 17 U.S.C. § 410(c) (2019) ("In any judicial proceedings the certificate of a registration made before or within five years after first publication of the work shall constitute prima facie evidence of the validity of the copyright and of the facts stated in the certificate. The evidentiary weight to be accorded the certificate of a registration made thereafter shall be within the discretion of the court.").

[80] *See id.* § 102(b) (2019); Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 556 (1985).

original work.[81] Because direct evidence can be difficult to obtain, plaintiffs often rely on indirect evidence established by showing that (a) the defendant had access to the copyright owner's work and (b) the two works are substantially similar.[82]

With respect to AI-generated works,[83] copyright infringement claims could arise in two neural-network-based machine-learning scenarios.[84] In the first infringement scenario, a claim may arise from an engineer's assembly of a digital corpus of training data, wherein a selection of copyrighted works is digitized and/or reproduced without authorization from the copyright owner,[85] thereby violating a copyright owner's exclusive right of reproduction.[86] Resolving this claim would involve a straightforward infringement analysis — if the plaintiff can show that the defendant made unauthorized copies of the plaintiff's work, the defendant will be liable for infringement unless otherwise excused.

In the second infringement scenario, a claim may arise when unauthorized intermediate copies of images are made during training in neural networks such as CNNs or the GAN's discriminative model.[87] How this infringement scenario might be resolved depends on which circuit the case is brought in. The Second and Fourth Circuits are likely to find that intermediate, ephemeral reproductions are not copies for purposes of infringement. But the Ninth, Eleventh, and D.C. Circuits

[81] N. Coast Indus. v. Jason Maxwell, Inc., 972 F.2d 1031, 1034 (9th Cir. 1992).

[82] *See, e.g.*, Sid & Marty Krofft Television Prods. v. McDonald's Corp., 562 F.2d 1157, 1162 (9th Cir. 1977); *see also infra* Part II.B.

[83] As previously described, using AI to generate artwork involves first assembling a large corpus of training data and then making intermediate copies of the data to train the AI program as its algorithm is refined over time. *See supra* notes 19–35 and accompanying text.

[84] To be clear, the infringement analysis is not of whether the output is infringing, but only whether using copyrighted inputs to create an AI-generated artwork is infringing.

[85] *See supra* Part I.A.

[86] *See* 17 U.S.C. § 106 (2019). Creating digital copies of copyrighted works through the use of photography or copying digital files would be considered an unauthorized reproduction. In *The Next Rembrandt* project, for example, the engineers digitized Rembrandt's portraits by taking high resolution photographs of the original paintings, effectively creating a photographic reproduction of the original painting that lacks independently created expression, and would be considered a "slavish cop[y]," and thus would be considered infringing activity had the Rembrandt portrait been protected by copyright. Bridgeman Art Library, Ltd. v. Corel Corp., 25 F. Supp. 2d 421 (S.D.N.Y. 1998), *modified on reconsideration by* 36 F. Supp. 2d 191, 197 (S.D.N.Y. 1999) (citing MELVILLE B. NIMMER, NIMMER ON COPYRIGHT § 22.08 (1963)).

[87] *See supra* note 32 and accompanying text.

would likely find that those exact same ephemeral reproductions are indeed infringing copies.

Section A first dispenses with the argument that creating intermediate reproductions in machine learning constitutes a non-infringing "operational" use of copyrighted works. Section B then examines whether the unauthorized intermediate copies from the second infringement scenario constitute infringing copies by reviewing the circuit split as to what constitutes an infringing copy. The Note will later argue that however the split is resolved, courts will likely reach the same conclusion that using copyrighted works in machine learning is fair use.[88]

A. Creating Intermediate Reproductions Is Not a Non-Infringing "Operational" Use of Copyrighted Works

The distinction that some scholars have drawn between a computer program's operational, creational, and input uses of copyright works does not accurately or sufficiently capture the use of such works to create AI-generated art. In *Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, the Second Circuit held that there is a distinction between a person who "volitionally operates the copying system to make the copy," and "issuing a command directly to a system, which automatically obeys commands and engages in no volitional conduct."[89] Professor Edward Lee has characterized these non-volitional uses of copyrighted works as intermediate "operational" uses, that is, uses that "occur during the operation of the technology [here, an AI program] *once it has already been created*."[90] Such operational use copies are made by computers, without volitional conduct, in order for the computer program to function *as designed*.[91] By contrast, a "creational use" means use of a copyrighted work "to *create* a technology."[92] Finally, "output uses" involve transferring some portion of the copyrighted input directly to the output.[93]

Under Professor Lee's framework, the three uses appear on a spectrum of potential infringement: "operational" and "creational" uses

---

[88] *See infra* Part III.

[89] Cartoon Network LP, LLLP v. CSC Holdings, Inc., 536 F.3d 121, 131 (2d Cir. 2008).

[90] Edward Lee, *Technological Fair Use*, 83 S. CAL. L. REV. 797, 843 (2010) (emphasis added).

[91] *See id.* at 843 n.231.

[92] *Id.* at 842 (emphasis added).

[93] *See id.* at 844.

are inherently non-infringing while "output" uses are inherently infringing. Applying this framework to AI programs, Professor Amanda Levendowski argues that training an AI program with copyrighted works is a "quintessential example of a 'purely operational' use under Professor Lee's framework."[94] However, using AI to generate artwork does not fall neatly into the operational use framework. Copyrighted works are not merely used during the operation of already-created technology; the AI program is not a static, fixed program that simply processes copyrighted works as data. Rather, the copyrighted works helps to build the AI program itself, which generates the artwork. As previously described, the AI program continues to evolve over time. The use of copyrighted works in AI also does not neatly fit as a "purely creational use"[95] because the AI program uses copies in its operations. The copies of the data, in theory, become part of "an output to end users."[96]

Because using copyrighted works in AI programs is not neatly excused by the operational-creational-output use framework, this Note next analyzes whether the creation of unauthorized intermediate copies in AI programs during machine learning is infringing.

B. There Is Uncertainty as to Whether Intermediate Reproductions in CNNs and the GAN's Discriminative Model Constitute Copies Under Copyright Law

This Section analyzes the claim that might arise in the second infringement scenario: whether intermediate reproductions of training data constitute unauthorized reproductions giving rise to infringement liability. As previously described in Part I.A *supra*, two of the primary neural networks used to train AI to generate images are the convolutional neural network and the generative adversarial network.[97] The CNN and the GAN's discriminative model both make use of intermediate reproductions of training data.[98]

When an engineer trains her AI program through the use of unauthorized intermediate copies of copyrighted works, the copyright owner might seek to bring an infringement claim against her. After establishing the first two elements of her claim, protectable subject

---

[94] Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579, 624 n.221 (2018).

[95] Lee, *supra* note 90, at 843 n.229.

[96] *Id.* at 843.

[97] *See supra* notes 23–25 and accompanying text.

[98] *See supra* notes 34–35 and accompanying text.

matter and the plaintiff's ownership of the copyrighted work, the plaintiff must then establish unauthorized copying of the protected work.[99] For an intermediate copy to be considered a copy under 17 U.S.C. § 101, it must be "'fixed' in a tangible medium of expression" in a state that is "sufficiently permanent or stable to permit it to be . . . reproduced . . . for a period of more than transitory duration."[100]

The question of whether intermediate reproductions of data during machine learning constitute unauthorized copies for the purposes of an infringement analysis has not been directly addressed by any U.S. court,[101] but there is a circuit split on similar questions involving older technologies — what is an unauthorized copy in an infringement analysis? This split reveals a lack of clarity as to whether the ephemeral intermediate reproductions used in CNNs and the GAN's discriminative model during machine learning constitute "copies" that might result in infringement liability.

1. Seventh, Ninth, and D.C. Circuits: Mere Downloading of an Unauthorized Reproduction onto a Computer Creates an Infringing Copy

In 1993, the Ninth Circuit decided *MAI Systems Corp. v. Peak Computer, Inc.*,[102] one of the first cases to address whether the embodiment of digitally received content was sufficiently fixed in a medium to constitute a copy.[103] In *MAI Systems*, plaintiff MAI was a computer and software manufacturer that licensed its software to its customers.[104] Defendant Peak was a third-party company that repaired and maintained MAI computers.[105] When Peak employees made repairs on a MAI computer, they downloaded MAI's software onto the client's computer's random access memory ("RAM"). While the MAI software

[99] *See* 17 U.S.C. § 102 (2019); *id.* § 411(a); Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991); *see also infra* Part II.B.2.

[100] 17 U.S.C. § 101 (2019).

[101] The following Westlaw search returned no case results as of May 16, 2020: artificial /s intelligence machine /s learning AND (intermedia! OR transit! OR interim) /10 (copy OR copies). Professor Edward Lee has argued that "[i]ntermediate copies used to create a new technology are distinguishable from interim copies of copyrighted works that are just used in the process of making the final copy." Lee, *supra* note 90, at 842 n.228. When the above Westlaw search was run without the terms artificial, intelligence, machine, and learning, the search returned 812 cases from all jurisdictions.

[102] 991 F.2d 511 (9th Cir. 1993).

[103] *See id.* at 517-19.

[104] *Id.* at 513, 517.

[105] *Id.* at 513.

licenses permitted MAI customers to use the software (therefore permitting MAI customers to download the software onto their computers' RAM), the license did not permit third parties like Peak to do the same.[106] MAI argued that by downloading its software onto the RAM, Peak created unauthorized copies and infringed MAI's copyright.[107]

The *MAI Systems* court affirmed the district court's holding that copying occurs when software "is transferred from a permanent storage device to a computer's RAM" or to a computer's central processing unit ("CPU")[108] and that in the absence of express permission by license or by ownership of the copyrighted software such acts constitute copyright infringement.[109] Thus, in the Ninth Circuit, the mere downloading of a copy of software onto a computer's RAM without appropriate permissions was enough to establish that an infringing copy was created.[110]

A few years later, the Seventh Circuit in *NLFC, Inc. v. Devcom Mid-America, Inc*. similarly held that downloading software onto a computer instantly created a copy under the Copyright Act.[111] The D.C. Circuit

[106] *Id.* at 517.

[107] *Id.* at 517-18.

[108] *Id.* at 518 (quoting 17 U.S.C. § 101) (explaining that when copyrighted software is downloaded onto a computer's RAM and can be diagnosed for problems, the representation of the software created in the RAM is "sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration."). Following *MAI Systems*, Congress added section 117(c) to the Copyright Act, which provided that creating a copy of a computer program *for repair or maintenance purposes* did not constitute infringement. *See* Title III of the Digital Millennium Copyright Act, Pub. L. 105-304, § 302(3) (1998) (emphasis added). Previously, section 117(a) permitted copying of software by the owner of the copyrighted software only if doing so was "an essential step in the utilization of the computer program" or was "for archival purposes only." 17 U.S.C. § 117(a) (Supp. 1988).

[109] *MAI Systems*, 991 F.2d at 518.

[110] *Id.* at 519 ("[I]t is generally accepted that the loading of software into a computer constitutes the creation of a copy under the Copyright Act." (citing Vault Corp. v. Quaid Software Ltd., 847 F.2d 255, 260 (5th Cir. 1988)); MELVILLE B. NIMMER & DAVID NIMMER, 2 NIMMER ON COPYRIGHT, § 8.08, at 8-105 (1983); NAT'L COMM'N ON NEW TECH. USES OF COPYRIGHTED WORKS, FINAL REPORT OF THE NATIONAL COMMISSION ON THE NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS 53, 61 (1981).

[111] NLFC, Inc. v. Devcom Mid-America, Inc., 45 F.3d 231, 235 (7th Cir. 1995) (citing MAI Sys. Corp. v. Peak Comput., Inc. and Vault Corp. v. Quaid Software Ltd. for the proposition that loading software onto a computer creates a copy under the Copyright Act).

held the same three years later in *Stenograph L.L.C. v. Bossard Associates, Inc.*[112]

Courts following the Seventh, Ninth, and D.C. Circuits' fixation standard will recognize that an infringing copy is created if a copyrighted image is simply downloaded onto a computer, no matter how long the image exists on the computer. With CNNs, a huge quantity of input data is temporarily stored in the computer's memory and potentially duplicated and deleted many times as the AI processes the underlying data patterns.[113] As a result, in these circuits, the creation of those ephemeral intermediate reproductions in CNNs would likely constitute copies for purposes of a copyright infringement claim.

2. Second and Fourth Circuits: More than Mere Downloading Is Required to Constitute a Copy

The Second and Fourth Circuits have required more than an ephemeral existence of a digital copy in order for a reproduction to constitute a copy under copyright law. In the Fourth Circuit case *CoStar Group, Inc. v. LoopNet, Inc.*, plaintiff CoStar owned and stored online a large database of images of homes, which it allowed real estate agents to use with proper attribution.[114] CoStar discovered that several of its copyrighted images were posted without permission on defendant LoopNet's real estate listing website by LoopNet's subscribers. Though LoopNet itself never posted copyrighted images, its regular business practice of reviewing images for compliance with its content policy required its employees to download all photos posted by subscribers. After downloading and reviewing the images of the homes, the employees deleted them. CoStar brought suit against LoopNet for infringement for the unauthorized downloading of the copyrighted photos.[115]

The Fourth Circuit panel affirmed the district court's grant of LoopNet's motion for summary judgment.[116] The court reasoned that LoopNet was not liable for direct infringement if its subscribers evoked in LoopNet a "temporary, automatic response to the user's request" — that is, to download the image onto LoopNet's RAM without regard to

[112] 144 F.3d 96, 102 (D.C. Cir. 1998) (affirming the *MAI Systems* holding that downloading results in "fixation" and consequently causes a "copy" to be made).

[113] *See* Xiaqing Li et al., *Performance Analysis of GPU-Based Convolutional Neural Networks*, 45 PROC. INT'L CONF. ON PARALLEL PROCESSING 67, 67 (2016).

[114] *See* CoStar Grp., Inc. v. LoopNet, Inc., 373 F.3d 544, 546-47 (4th Cir. 2004).

[115] *Id.*

[116] *Id.* at 546.

content.[117] The court explained that 17 U.S.C. § 106 contains an implicit requirement of "volitional conduct" by the alleged infringer, meaning that the defendant must know *that* she acted but does not have to know *whether* she was infringing.[118]

In the audiovisual context, the Second Circuit held in *Cartoon Network LP* that a brief embodiment of downloaded content for 1.2 seconds was not sufficiently fixed to constitute a copy.[119] The defendant in the case, Cablevision Systems Corporation ("Cablevision"), was a cable television systems operator that offered DVR services.[120] The DVR service allowed customers to record and play back copyrighted media from providers such as plaintiff Cartoon Network without obtaining an additional license.[121] The DVR functioned by redirecting part of the buffered media content to a Broadband Media Router ("BMR") located in the customer's home.[122] The buffered stream was stored on the BMR for 1.2 seconds before being routed back to Cablevision's storage facility, where the content was stored for future on-demand retrieval.[123] Cartoon Network argued that the 1.2-second storage of the content was unauthorized copying constituting infringement.[124]

The Second Circuit opinion emphasized that fixation requires not only that a work is embodied in a medium from where it can be later retrieved or reproduced, but also that it is embodied for more than a transitory duration.[125] The Second Circuit rejected the district court's reliance on *MAI Systems* to hold that *any* "transmission of information through a computer's . . . RAM . . . creates a 'copy' for purposes of the Copyright Act."[126] The court explained that although the *MAI Systems* court concluded the program was embodied on the RAM for more than a transitory duration, the reference to the "transitory duration" language was only cursory and thus could not be interpreted as holding that downloading a copy onto a computer's RAM "always results in

---

117 *Id.* at 550-51.

118 *Id.* at 551.

119 Cartoon Network LP, LLLP v. CSC Holdings, Inc., 536 F.3d 121, 129-30 (2d Cir. 2008).

120 *Id.* at 124.

121 *Id.* at 124, 127.

122 *Id.* at 124-25.

123 *Id.* at 125.

124 *Id.*

125 *Id.* at 127.

126 Twentieth Century Fox Film Corp. v. Cablevision Sys. Corp., 478 F. Supp. 2d 607, 621 (S.D.N.Y. 2007), *rev'd in part, vacated in part sub nom. Cartoon Network*, 536 F.3d at 121.

copying."[127] The Ninth Circuit would not have dismissed a statutory requirement after only a cursory mention.[128] Further, the court reasoned that if downloading *always* resulted in copying, the transitory duration requirement would be meaningless.[129] Ultimately, the court held the unauthorized presence of the content on the customer's BMR for 1.2 seconds was not fixed as a copy under 17 U.S.C. § 101.[130]

Courts following these circuits' fixation standard will likely look for evidence of the embodiment period before concluding that a copy has been created. The *CoStar Grp.* decision would be particularly instructive in the context of AI-generated art because the case involved downloading and reproducing digital images, which a CNN model and the GAN's discriminative model would also do.[131] It is thus unlikely under these circuits' fixation standard that CNNs' creation of ephemeral intermediate reproductions of images during training would constitute copies for the purpose of copyright infringement claims.

Although circuits differ in their standard for fixation, the more recent cases have tended to follow the Second and Fourth Circuits' 1.2-second standard for embodiment of a copy, rather than the Ninth Circuit's holding in *MAI Systems* that, generally, any downloading of software onto a computer creates a copy.[132] On the other hand, the *MAI Systems*' fixation standard has not been overruled and would still be followed by courts in the Ninth Circuit — a significant fact, given that this circuit

---

127 *Cartoon Network*, 536 F.3d at 127-28.

128 *See id.* at 128.

129 *See id.*

130 *Id.* at 129-30. The Second Circuit pointed to the U.S. Copyright Office's 2001 DMCA Report for further support that a duration requirement exists. *See id.* at 129. The report stated that a copy is not fixed if it is embodied "so fleetingly that it cannot be copied, perceived or communicated." *Id.* (quoting U.S. COPYRIGHT OFFICE, DMCA SECTION 104 REPORT 111 (2001)).

131 *See supra* Part I.B.

132 *Compare* MAI Sys. Corp. v. Peak Comput., Inc., 991 F.2d 511, 518-19 (9th Cir. 1993) (holding that the loading of copyrighted software creates a representation in the RAM that exists for more than transitory period and that "it is generally accepted that the loading of software into a computer constitutes the creation of a copy under the Copyright Act"), NLFC, Inc. v. Devcom Mid-Am., Inc. 45 F.3d 231, 235 (7th Cir. 1995) (citing *MAI Systems* and holding the same), *with* CoStar Grp., Inc. v. LoopNet, Inc., 373 F.3d 544, 551 (4th Cir. 2004) (finding that a computer's download of copyrighted material via an internet service provider as part of a "transmission function" does not constitute a sufficiently fixed copy lasting longer than a transitory period), *and Cartoon Network*, 536 F.3d at 129-30 (finding 1.2 seconds to be a transitory period and thus failing the duration requirement for embodiment).

frequently hears cases brought by and against many Silicon Valley and San Francisco technology companies.[133]

This Part has shown that whether copies created by AI machine learning programs last long enough to be fixed is uncertain. Some circuits have required more than an ephemeral existence of a digital copy in order for a reproduction to constitute a copy, while others have found that any downloaded copy is a sufficiently fixed copy under copyright law. Part III argues that regardless of the outcome of such an infringement analysis, the use of copyrighted works in training an AI program is fair use.

III. USING COPYRIGHTED WORKS TO TRAIN AI IS FAIR USE

Part II described the existing circuit split as to what constitutes an infringing copy. The Seventh, Ninth, and D.C. Circuits have held that any intermediate reproduction on a computer is sufficiently fixed to constitute an infringing copy. The Second and Fourth Circuits have held that only intermediate reproductions that exist for some period of time are sufficiently fixed to constitute infringing copies. This Part argues that, regardless of how the split is resolved, creating intermediate reproductions of copyrighted works to train AI programs is a fair use.

The central purpose of the fair use doctrine is to "guarantee . . . breathing space within the confines of copyright."[134] The fair use doctrine expressly classifies certain uses — such as criticism, commentary, news reporting, teaching, scholarship, and research — as non-infringing, even if they would otherwise violate statutory exclusive rights.[135] However, this list does not exhaust all possible fair uses. The Supreme Court has held that analyzing fair use requires a case-by-case analysis, rather than application of bright-line rules.[136] Fair use is analyzed under four factors: (1) the purpose and character of the use, (2) the nature of the copyrighted work, (3) the amount and substantiality of the portion used in relation to the copyrighted work as

---

[133] *See, e.g.*, Vernor v. Autodesk, Inc., 621 F.3d 1102, 1110-11 (9th Cir. 2010); MDY Indus., LLC v. Blizzard Entm't, Inc., 629 F.3d 928, 938 (9th Cir. 2010), *as amended on denial of reh'g* (Feb. 17, 2011), *opinion amended and superseded on denial of reh'g*, No. 09-15932, 2011 WL 538748 (9th Cir. Feb. 17, 2011) (citing *MAI Systems* and finding that when a computer game is played, the game's software is copied onto the computer's RAM which "potentially infringes" the copyright unless the player is a licensee using the software within the scope of the license or owns a copy of the software).

[134] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994).

[135] *See* 17 U.S.C. § 107 (2019); Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 95-96 (2d Cir. 2014).

[136] *See Campbell*, 510 U.S. at 577.

a whole, and (4) the effect of the use upon the potential market for or value of the protected work.[137] None of these fair use factors is dispositive, rather, they are weighed holistically.[138] The fair use doctrine is an equitable rule of reason, and as such, courts may "adapt the doctrine to particular situations on a case-by-case basis."[139]

The Second Circuit has held that large-scale digitization for the purpose of data manipulation is fair use because it differs from the original works' purpose to convey aesthetic, informational, or other expressive content. Indeed, the Second Circuit has done so in two infringement cases involving the digitization of hard-copy books in the research-institution[140] and commercial[141] contexts. Still, scholars and judges alike opine that such cases push the boundaries of the fair use doctrine.[142] Notwithstanding these arguments, both (a) the digitization of copyrighted works to create a corpus of training data for machine learning, and (b) the creation of intermediate copies during machine learning are likely to fall within the boundaries of the fair use doctrine. The Sections that follow analyze how.

A. Purpose and Character of the Use

The first and fourth fair use factors are generally regarded as the most influential in court decisions.[143] The first factor, the purpose and character of the use, involves evaluating two connected issues: whether

---

[137] *See* 17 U.S.C. § 107 (2019).

[138] *See Harper & Row Publishers, Inc.*, 471 U.S. at 549.

[139] Hustler Magazine, Inc. v. Moral Majority, Inc., 796 F.2d 1148, 1151 n.4 (9th Cir. 1986) (quoting H.R. REP. NO. 94-1476, at 65-66, *reprinted in* 1976 U.S.C.C.A.N. 5659, 5679-80); *see also* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 577 (1994) (stating that fair use cannot "be simplified with bright-line rules, for the statute, like the doctrine it recognizes, calls for case-by-case analysis").

[140] *See* Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 103 (2d Cir. 2014).

[141] *See* Authors Guild v. Google, Inc., 804 F.3d 202, 206-207, 229-230 (2d Cir. 2015) (holding that Google's digitization of tens of millions of books submitted by major U.S. libraries for the purpose of making the works digitally searchable online, i.e., allowing users to search whether a book contained a certain term without making the entire book accessible, was fair use and did not constitute infringement).

[142] For example, the Second Circuit recently held that viewing up to ten minutes of copyrighted news clips assembled by the defendant in a searchable database did not constitute fair use. *See* Fox News Network, LLC v. TVEyes, Inc., 883 F.3d 169, 178-81 (2d Cir. 2018) (discussing the fair use defense).

[143] *See* Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005*, 156 U. PA. L. REV. 549, 586 (2008) (using correlation and regression analyses to examine the four fair use factors and concluding that "the outcome of the fourth factor appears to drive the outcome of the test, and that the outcome of the first factor also appears to be highly influential").

the allegedly infringing use is commercial in nature and whether the use is transformative.[144] A finding of transformative use strongly weighs in favor of fair use, while a finding that the use is commercial in nature weighs against it.[145] Further, the more transformative a work is, the less significant the use's commercial nature will be to the court's analysis.[146]

Transformation involves "altering the original with new expression, meaning, or message."[147] Works are considered transformative only when the copyrighted works are changed or used "in a different context such that the . . . work is transformed into a new creation."[148] Mere repackaging of works into different formats is not transformative.[149] Rather, transformation requires creating within the allegedly infringing work new or different purposes or functions as compared to those of the original.[150]

---

[144] *See* Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 562-63 (1985).

[145] *See id.*

[146] *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 569 (1994).

[147] *Id.*

[148] Oracle Am., Inc. v. Google LLC, 886 F.3d 1179, 1202 (Fed. Cir. 2018).

[149] *See, e.g.*, Castle Rock Entm't, Inc. v. Carol Publ'g. Grp., Inc., 150 F.3d 132, 142 (2d Cir. 1998) (finding that the purpose of the Seinfeld Aptitude Test was merely to "repackage *Seinfeld* to entertain *Seinfeld* viewers" and thus "[a]ny transformative purpose possessed by *The SAT* is slight to non-existent"); Soc'y of the Holy Transfiguration Monastery, Inc. v. Archbishop Gregory of Denver, 685 F. Supp. 2d 217, 227 (D. Mass. 2010), *aff'd*, 689 F.3d 29 (1st Cir. 2012) (finding that repackaging a copyrighted work in new format was not transformative "when the result is simply a mirror image reflected on a new mirror").

[150] Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 96 (2d Cir. 2014); *see also* Blanch v. Koons, 467 F.3d 244, 252-53 (2d Cir. 2006) (citing Bill Graham Archives LLC. v. Dorling Kindersley Ltd., 448 F.3d 605, 609 (2d Cir. 2006)) (holding that for use to be transformative, the purpose of the use within the new work must differ from the original author's purpose in creating the copyrighted work). An example of a different purpose may be digitizing copyrighted works to enable online search functionality. *See, e.g.*, Authors Guild v. Google, Inc., 804 F.3d 202, 216-17 (2d Cir. 2015) (finding Google's creation of digital copies of copyrighted works served the "purpose of enabling a search for identification of books" and thus "involves a highly transformative purpose"); Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007) (finding Google's use of thumbnails of copyrighted images was transformative because instead of serving "an entertainment, aesthetic, or informative function," the thumbnails served as "pointer[s] directing a user to a source of information"); *see also, e.g.*, Kelly v. Arriba Soft, 336 F.3d 811, 818, 822 (9th Cir. 2003). Another example of a different purpose may be using the copyrighted work to provide historical context for a given subject matter. *See, e.g.*, Bill Graham Archives, LLC v. Dorling Kindersley Ltd., 386 F. Supp. 2d 324 (S.D.N.Y. 2005), *aff'd*, 448 F.3d 605 (2d Cir. 2006) (finding that using smaller sizes of copyrighted poster images in a biographical book was "not used to directly attract sales of the book," but rather to provide historical context).

Several circuit court cases provide relevant guidance in analyzing whether creating and manipulating digital reproductions of works is transformative. In *Authors Guild, Inc. v. HathiTrust (HathiTrust)*, an organization called HathiTrust electronically scanned physical books from several research universities and made the digitized collection of these books accessible to any of HathiTrust's approximately eighty member research institutions.[151] The HathiTrust library permitted three uses of this digitized collection, one of which allowed the public to generally access the digitized books, that is, to search for specific terms within them, but not to have access to the entire book.[152] The public could search for terms across all digital copies and receive search results comprised of the page location of the term and the number of times the term appeared on the page.[153]

One of the questions before the Second Circuit was whether the creation of a full-text searchable database, which necessarily required "creat[ing] digital copies of the entire books," constituted fair use.[154] The court answered in the affirmative.[155] In analyzing the first fair use factor, whether the purpose and character of the use is commercial or transformative, the court concluded that creating a full-text searchable database is a "quintessentially transformative use."[156] The court reasoned that HathiTrust's granting of public access to the digitized books for search purposes was fair because "the result of a word search is different in purpose, character, expression, meaning, and message" from the original book.[157] According to the Second Circuit, digitizing books was "quintessentially transformative," and the court resolved the first factor in favor of fair use.

The Second Circuit reached a similar conclusion in the context of artwork in *Blanch v. Koons*. The defendant-artist Koons made a digital copy of a single copyrighted photograph of a woman's legs and sandals created by the plaintiff-artist, Blanch.[158] Koons reproduced a digitized

---

[151] *See HathiTrust*, 755 F.3d at 90.

[152] *See id.* There were two other uses of full-text searchable database of digitized books. One was to enable individuals with print-disabilities could access the full texts of the copyrighted works. *See id.* The other use was that the HathiTrust library permitted member institutions that owned an original copy of a work to create a replacement copy if their original copy was stolen, lost, or destroyed and a replacement was not otherwise obtainable at a "fair" price. *Id.* at 92.

[153] *See id.* at 91.

[154] *Id.* at 97.

[155] *Id.*

[156] *Id.*

[157] *Id.*

[158] *See* Blanch v. Koons, 467 F.3d 244, 247-49 (2d Cir. 2006).

version of Blanch's photograph in his own painting, onto which he added three other women's legs, painted as if they were dangling from the top of the painting.[159] The Second Circuit found that in using Blanch's photograph in his painting, Koons had a "sharply different objective[]" than Blanch had when she created her photograph.[160] Specifically, Koons "want[ed] the viewer to think about his/her personal experience with these objects, products, and images and at the same time gain new insight into how these affect our lives," while Blanch "wanted to show some sort of erotic sense[;] . . . to get . . . more of a sexuality to the photographs."[161] The court concluded that Koons' use of Blanch's photograph was "transformative" and "strongly" weighed in favor of fair use.[162]

In *Authors Guild, Inc. v. Google, Inc.*,[163] a case involving a digital index of over 20 million books in the Google Books search database, including those whose copyrights were owned by Authors Guild, the Second Circuit held that even in the commercial context, the database did not infringe Authors Guild's copyrights. The Second Circuit affirmed its earlier views in *HathiTrust* and reasoned that the creation of complete digital copies of copyrighted works had the purpose of providing a search function,[164] which was transformative fair use that "served a different function from the original."[165]

Following the Ninth Circuit, the Second Circuit in *NXIVM Corp. v. Ross Institute*[166] and the Federal Circuit in *Oracle America, Inc. v. Google LLC*[167] have held that an alleged infringer's bad faith or "unclean hands" does not preclude the fair use defense.[168] Thus, even if the programmer wished to build an AI program that could create works that are very similar to other copyrighted works, the programmer would not be liable for infringement if the fair use defense otherwise applies.

Applying this reasoning to the context of AI-generated artwork, an engineer could similarly digitize copyrighted artworks to assemble a

159 *See id.* at 247-48.

160 *Id.* at 252.

161 *Id.*

162 *Id.* at 253, 256.

163 Authors Guild v. Google, Inc., 804 F.3d 202, 216-17 (2d Cir. 2015).

164 *See Id.* at 216-17.

165 *Id.* (quoting Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 97 (2014) (citing Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007))); *see also* A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630, 639-40 (4th Cir. 2009); Kelly v. Arriba Soft Corp., 336 F.3d 811, 819 (9th Cir. 2003).

166 NXIVM Corp. v. Ross Inst., 364 F.3d 471 (2d Cir. 2004).

167 886 F.3d 1179 (Fed. Cir. 2018).

168 *Id.* at 1203-04; *NXIVM*, 364 F.3d at 479.

corpus of training data for her AI program. This use is highly transformative because the original purpose of an author or artist in creating expressive works is to communicate through artistic expression, earn a living, and/or practice their craft. By contrast, the engineer converts such expressive works into training data to allow an AI program to refine its algorithm by minimizing the error between generated works and the ideal output, which is informed by the training data, as previously described.[169] Since digitizing copyrighted images to train an AI program is a different purpose than the purpose underlying an artist's creation of a work of art, digitization to train an AI program would likely be considered transformative. The first factor therefore weighs in favor of fair use.

B. The Nature of the Copyrighted Work

The second fair use factor, the nature of the copyrighted work, considers whether the copyrighted work is "of the creative or instructive type that the copyright laws value and seek to foster."[170] The more creative, fictional, or fantastical (as opposed to factual) the copyrighted work is, the less likely a court will find in favor of fair use.[171]

In *HathiTrust*, the second factor, the nature of the copyrighted work, received little analysis but was also resolved in favor of fair use.[172] The Second Circuit acknowledged that this factor may be of limited use where the work being used for a transformative purpose is "creative."[173] Similarly, in *Blanch*, the court only briefly discussed the second fair use factor. The court examined two distinctions: (a) whether the allegedly infringing work was creative and expressive in nature, which would weigh in favor of fair use, as opposed to being more "factual or informational," and (b) whether the allegedly copied work is

---

169 *See supra* notes 13–40 and accompanying text.

170 *HathiTrust*, 755 F.3d at 96 (quoting Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1117 (1990)).

171 *See* Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1117 (1990) ("[T]he second factor should favor the original creator more heavily in the case of a work (including superseded drafts) created for publication, than in the case of a document written for reasons having nothing to do with the objectives of copyright law."). *But see* Stewart v. Abend, 495 U.S. 207, 237 (1990); Micro Star v. Formgen Inc., 154 F.3d 1107, 1113 (9th Cir. 1998).

172 *See HathiTrust*, 755 F.3d at 98, 103.

173 *Id.* at 98 (quoting Bill Graham Archives LLC v. Dorling Kindersley Ltd., 448 F.3d 605, 612 (2d Cir. 2006)).

published.[174] Because Blanch's original photograph was publicly accessible, this factor also weighed in favor of a finding of fair use.

In the context of AI-generated artwork, the images that a defendant-engineer might copy are published because they have been posted on the internet and would thus be publicly accessible, enabling the engineer to download the images to assemble her corpus of training data. Therefore, this factor would also likely weigh in favor of fair use.

C. Amount and Substantiality of the Portion Copied

The third fair use factor is the amount and substantiality of the portion copied or used in relation to the copyrighted work as a whole.[175] With respect to this factor, the *HathiTrust* court held that retaining copies of the digitized works as images and as text-only files was not excessive because it was reasonably necessary for HathiTrust "to make use of the entirety of the works in order to enable the full-text search function."[176] The *Blanch* court emphasized that the third factor analysis focuses on whether the amount copied is "reasonable in relation to the purpose of the copying."[177] The question is whether the alleged infringer copied more than what was "justified" to achieve his stated purpose for copying.[178] The court concluded that the amount copied was justified, given how different Blanch's purpose in taking the original photograph was from Koon's purpose in copying it in his own painting.[179]

In the context of an AI-generated artwork, a court would likely similarly find that even though the images were copied wholesale to form the training data, the purpose of using the copied images to train the AI is so different from the copyright owner's original purpose in creating the copyrighted work that such reproductions would be justified.

D. Effect Upon the Potential Market for or Value of the Copyrighted Work

The fourth factor, the effect upon the potential market for or value of the copyrighted work, imposes the condition that fair use should not

---

[174] Blanch v. Koons, 467 F.3d 244, 256 (2d Cir. 2006).

[175] *See supra* note 137 and accompanying text.

[176] *HathiTrust*, 755 F.3d at 98.

[177] *Blanch*, 467 F.3d at 257 (citing Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 586 (1994) (quoting Folsom v. Marsh, 9 F. Cas. 342, 348 (1841))).

[178] *Id.*

[179] *See id.* at 257-58.

excessively damage the market for the original work by serving as a viable substitute for it.[180] While courts emphasize the importance of transformative use,[181] the fourth factor is still considered the primary driver of the test.[182]

The aim of the fourth fair use factor is to "assess the impact of the use on the traditional market for the copyrighted work."[183] The impact on the traditional market must stem from the fact that the result of the allegedly infringing use serves as a market substitute for the original work.[184] When "two works usually serve different market functions," they are likely not substitutes.[185] Further, the owner must show that if the challenged use becomes widespread, it would have an adverse effect on the potential market for her work.[186] Thus, in the context of AI-generated artwork, a court will ask whether the AI-generated work would be a substitute for the original artist's work, either now or in the future.

In *HathiTrust* and *Blanch*, the fourth factor was resolved in favor of fair use.[187] The HathiTrust libraries argued that full-text search would not harm any existing or potential market. The Second Circuit found that because the sufficiency of the security measures undertaken by the libraries was unrebutted, there was a low likelihood that there would be a "public release of the specific copyrighted works," and that the full-text search function did not serve as a substitute for the books being searched.[188] Therefore, the court concluded, there was likely no

[180] *See HathiTrust*, 755 F.3d at 95-96.

[181] *See Campbell*, 510 U.S. at 579 (explaining that transformative use "generally further[s]" copyright's goal of "promot[ing] science and the arts," and thus transformative works "lie at the heart of the fair use doctrine's guarantee of breathing space within the confines of copyright (quoting Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 455 n.40, 477-80 (Blackmun, J., dissenting))); *see also* Oracle Am., Inc. v. Google LLC, 886 F.3d 1179, 1198 (2018).

[182] *See* Beebe, *supra* note 143, at 586; *see* Fox News Network LLC v. Tveyes, Inc., 883 F.3d 169, 176 (2d Cir. 2018) (citing Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 566 (1985)).

[183] *HathiTrust*, 755 F.3d at 96 (quoting *Harper & Row*, 471 U.S. at 566).

[184] *See Campbell*, 510 U.S. at 570-71.

[185] *Id.* at 570.

[186] *See* Hustler Magazine, Inc. v. Moral Majority, Inc., 796 F.2d 118, 1156 (1986) (finding that where the work resulting from the defendant's use "could not have diminished any potential sales, interfered with the marketability of the parody or fulfilled the demand for the original work," the defendant will have "rebutted any presumption of unfair exploitation"); *see also Harper & Row*, 471 U.S. at 566-67; *Sony*, 464 U.S. at 451.

[187] *See HathiTrust*, 755 F.3d at 99-100; Blanch v. Koons, 467 F.3d 244, 258 (2006).

[188] *HathiTrust*, 755 F.3d at 100.

impending harm to the market for the plaintiffs' works.[189] The *Blanch* court similarly resolved this fourth factor in favor of the defendant Koons, holding that because Blanch acknowledged that Koons' use of her photograph "did not cause any harm to her career" and the value of her photograph "did not decrease," Koons's use did not have a negative effect on the market for Blanch's work.[190]

The fear that AI-generated artworks will harm original artistry should not be quickly dismissed. After all, artists spend years refining their craft to create unique aesthetics and expressions that channel their emotions and lived experiences. In fact, the portraits that inspired one of the first widely publicized creations of AI-generated art, the *Rembrandt Project*, were created during the most difficult, final years of Rembrandt's life.[191] It was during this tragedy-filled period — during which he first lost his wife, two of his three sons, and then later, his lover and last son — that Rembrandt adopted "unorthodox technique[s]" and painted his most "experimental and exuberantly creative" works.[192] Also during this time, a painting he had been commissioned to produce that hung in Amsterdam's Town Hall was removed and replaced,[193] and he had to sell his drawings, paintings, furniture, and house, and lived the rest of his life in bankruptcy.[194]

Art historians and scholars evaluating the corpus of Rembrandt's many self-portraits propose that his prolific creation functioned to pay homage to "art in the eyes of the art-loving collector."[195] Over the course of his life, Rembrandt's self-portraits reflected his aging, his self-reflections and self-observations, and the personal experiences he lived through.[196] In particular, scholars note that the massive personal and financial losses that occurred during his final years were accompanied

189 *See id.* at 101.

190 *Blanch*, 467 F.3d at 258.

191 *See* Morris, *supra* note 2.

192 *Id.* In 1642, his wife Saskia van Uylenburgh died at the age of thirty-one, shortly after the birth of one of his three sons, Titus, the only child to survive into adulthood. *See* Marcus & Clarfield, *supra* note 1, at 27; J.W., *Rembrandt: The Late Works: Portrait of the Artist as an Old Man*, ECONOMIST (Oct. 17, 2014), https://www.economist.com/prospero/2014/10/17/portrait-of-the-artist-as-an-old-man [https://perma.cc/3EVK-R7Z3]. Between 1662 and 1669, both his lover Hendrickye Stoffels and son Titus died. *See* Morris, *supra* note 2.

193 Morris, *supra* note 2.

194 *See* Marcus & Clarfield, *supra* note 1, at 27.

195 *Id.* at 29-30.

196 *See id.* ("Many scholars have analyzed Rembrandt's series of self-portraits as reflecting a combination of the aging process along with Rembrandt's life events."); *id.* at 31 (citing JAKOB ROSENBERG, REMBRANDT: LIFE AND WORK (1964) for assertion on Rembrandt's self-observations).

by a distinct change in his painting style,[197] specifically, a shift toward "self-acceptance, integrity, and spiritual illumination."[198] This deviated from the infusions of lightness and youthful "arrogance" that characterize the self-portraits of his younger years.[199] It was Rembrandt's last decade of life, when he experienced the most tragedy, that some scholars regard as the period responsible for the master's most "impressive" works.[200] There seems to be something profoundly disturbing with allowing a lifeless computer to potentially disturb the market for artworks born of human experiences, including tragedy and suffering.

However, AI-generated works' disturbance of the market for original artists' works would likely be small. A consumer who wishes to buy an original artwork from a particular artist will likely not accept the work of another artist as a substitute, even though the substitute artist's work may look similar to the original artist's. Especially in the context of the market for artworks by reowned artists, consumers are unlikely to simply buy a print of the original artwork.[201] The market that this consumer participates in is mostly comprised of wealthy individuals, institutions, businesses, or collectors because original pieces are so expensive.[202] Even if we were to consider artworks by artists that are not as well-known as Rembrandt, consumers of art are likely to value owning an original work of art rather than a reprint or a computer-generated piece that resembles or recalls the original. The market for

197 *See id.* at 30 ("Neiderland explains that as a result of Rembrandt's personal and financial losses in later life, the artist altered the scope of his art and concentrated on portraying man's spiritual life." (citing W. G. Neiderland, *Psychoanalytic Concepts of Creativity and Aging: Psychoanalytic Approaches to Creativity*, 6 J. GERIATRIC PSYCHIATRY 160 (1973))).

198 *Id.* at 33.

199 *Id.* at 30 (stating that the portraits "move from the jauntiness, vitality and even arrogance of youth and early adulthood, to the quiet, introspective and intensely spiritual experiences of old age" (quoting G.E. Berg & Sally Gadow, *Toward More Human Meaning of Aging: Ideals and Images From Philosophy and Art*, *in* AGING AND THE ELDERLY: HUMANISTIC PERSPECTIVES IN GERONTOLOGY 83-92 (Stuart F. Spicker et al. eds., 1978))).

200 *Id.* at 33 (citing Gisela Labouvie-Vief, *Psychological Transformations and Late-Life Creativity*, 11 J. MUSEUMS ART & ARCHEOLOGY U. MICH. 71, 71 (1996)); *see also* J.W., *supra* note 192 ("During the often sad and sometimes humiliating years this show covers — from 1653 to his death in 1669 — his painting became dark and sombre. Its mood was frequently introspective; his brush-strokes were often impressionistic and his deployment of oil was, deliberately, the opposite of precise.").

201 *See* Nikki Martinez, *The Art of Buying Art*, HUFFPOST (Sept. 12, 2017, 12:01 AM), https://www.huffingtonpost.com/entry/the-art-of-buying-art_us_59b3603ae4b0bef3378ce048 [https://perma.cc/ENB4-UZGV].

202 *See id.*

AI-generated artworks therefore likely does not overlap with the market for original artwork.

While there is a counterargument that some price-sensitive consumers are happy with a less expensive print of an original,[203] it is more likely that both types of consumers consider AI-generated art to be an entirely new category of art. Although the AI community has known about AI applications in art for many years, AI-generated art is only beginning to enter the general public's consciousness. In 2016, a Turkish artist sold an AI-generated artwork entitled *GCHQ* for $8000 at a Google charity auction in San Francisco.[204] On October 25, 2018, for the first time in history, an AI-generated work created by the art collective Obvious was auctioned at Christie's in New York City.[205] The work, entitled *Portrait of Edward Bellamy*, sold for $432,500, about forty-five times higher than what Christie's initially estimated.[206]

---

[203] *See id.*

[204] *See* Sam Gaskin, *When Art Created by Artificial Intelligence Sells, Who Gets Paid?* , ARTSY (Sept. 17, 2018, 11:45 AM), https://www.artsy.net/article/artsy-editorial-art-created-artificial-intelligence-sells-paid [https://perma.cc/3GCD-YDVQ].

[205] Many news outlets covered the first sale of an AI-generated artwork at the renowned auction house. *See, e.g.*, Cohn, *supra* note 49; *This AI-Generated Portrait Just Sold for a Stunning $432,500*, CBS NEWS (Oct. 26, 2018, 8:33 AM), https://www.cbsnews.com/news/ai-generated-portrait-sells-for-stunning-432500-portrait-of-edmond-de-belamy/ [https://perma.cc/3NPB-9P3T]; William Falcon, *What Happens Now That an AI-Generated Painting Sold for $432,500?*, FORBES (Oct. 25, 2018, 08:29 PM), https://www.forbes.com/sites/williamfalcon/2018/10/25/what-happens-now-that-an-ai-generated-painting-sold-for-432500/#5b47810ca41c [https://perma.cc/4KQ7-KQCY]; Melanie Gerlis, *Old Art Responds to New Media*, FIN. TIMES (Nov. 1, 2018), https://www.ft.com/content/f407191a-dcf5-11e8-b173-ebef6ab1374a [https://perma.cc/WUS5-EGRZ]; Vanessa Romo, *A.I. Produced 'Portrait' Will Go Up For Auction At Christie's*, NPR (Oct. 22, 2018, 10:08 PM), https://www.npr.org/2018/10/22/659680894/a-i-produced-portrait-will-go-up-for-auction-at-christie-s [https://perma.cc/8Y9Z-VBXV].

[206] *See Is Artificial Intelligence Set to Become Art's Next Medium?*, CHRISTIE'S (Dec. 12, 2018), https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx [https://perma.cc/C5Q5-SVUA]. This auction was not without controversy. The Obvious group appeared to have used open-source GAN code written by Robbie Barrat, who had shared his code on the website GitHub and gave assistance to the group as they sought to build their GAN network. *See* James Vincent, *How Three French Students Used Borrowed Code to Put the First AI Portrait in Christie's*, VERGE (Oct. 23, 2018, 9:34 AM), https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robbie-barrat-gans [https://perma.cc/N84L-7DPP]. When that much economic value can be instantaneously generated by a third party that spent little time developing artistic sensibilities, people start paying attention. While the auction was ongoing, Barrat expressed surprise when he learned his open source code, which he had posted on the website GitHub, was used by the Obvious group to pursue a high-priced auction sale. *See* robbiebarrat, *art-DCGAN*, GITHUB, https://github.com/robbiebarrat/art-DCGAN (last visited Jan. 31, 2020) [https://perma.

These AI-generated artworks likely sell for such high prices primarily because of the novelty of their source — computers — and not because they resemble the paintings used as input data for the AI program. Therefore, while in certain cases the market for AI works may overlap with that of the original artist, consumers will usually regard AI-generated art as belonging to a different category of art. There might thus only be a small impact on the market for human-created artwork.

Further, using Rembrandt's painful lived experiences that inspired his brilliant paintings to justify shrinking the market for AI-generated art that resembles his works requires relying on the moral rights theory of copyright law, which as discussed above, is a theory credited neither by the Constitution nor the U.S. Supreme Court.[207] The constitutional purpose of copyright law is to incentivize the production of works regardless of who the author is — whether machine or human. Using AI to generate artworks thus aligns with the utilitarian theory, which prevails in U.S. copyright doctrine. Consequently, the fourth fair use factor will likely also be resolved in favor of fair use.

CONCLUSION

While there are some calls for the Copyright Act to be amended,[208] there may be no need to expressly address copyright issues that arise from the use of AI to generate artworks. Using AI to generate artwork marks a promising technological advance, but there have been real concerns that copyright law's uncertain application to these works casts a pall over their future.[209] Further, although regulating copyright-related activities and potentially infringing uses is difficult, advances in technology make enforcement of copyright law on digital platforms

cc/5TCW-EDGG];Vincent, *supra*. Barrat wrote on Twitter, "Does anyone else care about this? Am I crazy for thinking that they really just used my network and are selling the results?" Robbie Barrat (@videodrome), TWITTER (Oct. 24, 2018, 7:31 AM), https://twitter.com/videodrome/status/1055285640420483073?lang=en [https://perma.cc/K5V6-PSPS]. He explained that he had trained the GAN and shared the code online more than a year ago. *See id.* If a programmer who freely shared his code online and who actively helped third parties utilize the code felt surprised and dismayed that his work was used to generate huge sums of money, it is conceivable that an artist would feel similarly if she discovered that others had profited from her aesthetic.

[207] *See supra* notes 6–7, 65 and accompanying text.

[208] *See supra* note 10 and accompanying text.

[209] *See supra* Part I.B (discussing that the Copyright Act does not expressly contemplate authorship of computer-generated works and the decades-long lack of consensus among copyright scholars about which entity should own the copyright in computer-generated works).

increasingly practicable.[210] Nonetheless, this Note has shown that using copyrighted works to train AI programs that generate art might be permitted in some courts, or that such uses would be considered fair use. Expanding the permissible uses of copyrighted works in machine learning helps engineers continue making advances in computer science and technology, an outcome that is not only socially desirable but constitutionally approved.[211]

[210] *See* JOHN TEHRANIAN, INFRINGEMENT NATION xx-xxi (2011).

[211] *See supra* notes 6, 65–67 and accompanying text.