RUTGERS BUSINESS LAW REVIEW

VOLUME 18 SPRING 2023 NUMBER 2

ARTICLES

NOTES

BUILDING AND USING GENERATIVE MODELS UNDER US COPYRIGHT LAW

Van Lindberg[1]

Advancements in artificial intelligence (AI) have made it possible to generate new text, code, pictures, and music.[2] Users have described the results as "pure magic."[3] The popularity and capability of these tools have prompted reams of speculation about how they will transform society.

These new tools are powered by machine learning (ML) techniques. Machines "learn" by digesting millions of example sentences, software functions, pictures, or songs. The tools then use the accumulated "knowledge" to generate new works. For copyright

---

[1] Author is a Partner at Taylor English Duma LLP. "Building and Using Generative Models Under US Copyright Law," Volume 18 Rutgers Bus. L.R. 1 (2023). The author wishes to thank Garrett Hostetter, BYU Law School class of 2023 for his feedback and research assistance.

[2] Bernard Marr, *Beyond ChatGPT: 14 Mind-Blowing AI Tools Everyone Should Be Trying Out Now*, FORBES (Feb. 28, 2023, 2:31 AM), https://www.forbes.com/sites/bernardmarr/2023/02/28/beyond-chatgpt-14-mind-blowing-ai-tools-everyone-should-be-trying-out-now/?sh=305066b97a1b (previewing AI tools for text generation (ChatGPT and Murf), art creation (DALL-E 2 and Stable Diffusion2), video creation (Lumen5, Gen1, and Deep Nostalgia), music generation (Soundraw and Podcastle), and other creative works (Looka)); KELVIN CHAN AP, *What can ChatGPT maker's new AI model GPT-4 do?*, ABC NEWS (Mar. 15, 2023, 9:35 AM), https://abcnews.go.com/International/wireStory/chatgpt-makers-new-ai-model-gpt-4-97881867 (previewing Chat GPT-4); Christopher Tozzi, *GitHub Copilot vs. Amazon CodeWhisperer*, ITPRO TODAY (Jan. 26, 2023), https://www.itprotoday.com/development-techniques-and-management/github-copilot-vs-amazon-codewhisperer-what-developers-need (comparing two software AI tools); Alex Hughes, *Midjourney: The gothic AI image generator challenging the art industry*, SCIENCE FOCUS (Feb. 7, 2023, 4:40PM), https://www.sciencefocus.com/future-technology/midjourney/ (describing a popular AI art generator).

[3] Jim Carroll, "*The Nearest Thing To Magic Can Now Emerge In Moments!"*, JIM CARROLL (Sept. 12, 2022), https://jimcarroll.com/2022/09/daily-inspiration-the-nearest-thing-to-magic-can-now-emerge-in-moments/; Yanir Seroussi, *ChatGPT is transformative AI*, Yanir Seroussi (Dec. 11, 2022), https://yanirseroussi.com/2022/12/11/chatgpt-is-transformative-ai/; User (@u/devdef), REDDIT (Jan. 5, 2021, 7:55 PM), https://www.reddit.com/r/computervision/comments/krennw/openai_text2image_model_is_pure_magic/; All About AI, *Midjourney V4 - Image to Image: Pure Magic*, YOUTUBE (Nov. 8, 2022), https://www.youtube.com/watch?v=tQ_-NADMGcg.

purposes, much of the ML training is accomplished using copies of millions of different works as inputs to the learning process. Almost all of these works are copyrighted.

Machine learning is not new. Its theoretical foundations were established in the 1960s and working systems were created in the 1980s and 1990s. What is new is *scale* and *quality*. Scientists have harnessed the increasing capability of computers and the explosion of digital content to create software programs that rival humans in the ability to generate pictures, text, or music. Earlier AI technologies produced outputs that were clearly mechanical. In contrast, these new tools appear to be so smart that some people incorrectly describe what they do as if they were human and have human intentions and motivations.

ML applications are meant to produce wholly new outputs—but sometimes the applications reproduce fragments, or even whole copies, of works used in training. This raises important questions about copyright infringement. Further, even if outright copying does not occur, these ML applications can generate works that recall the style of specific authors and artists, causing worries that ML may outcompete and replace human creators.

Predictably, this new use of copyrighted material has already prompted lawsuits.[4] Whenever new uses of copyrighted works emerge, fights for control follow. Copyright holders unsurprisingly want to be paid for the use of their works to build these ML models.

This article addresses the legal issues associated with building and using ML models from a technology-first perspective. It explains machine learning models, how they are trained, and how they generate new works. It then analyzes the applicable law,

[4] *See* Doe v. GitHub Inc., No. 3:22-CV-06823 (N.D. Cal. filed Nov. 3, 2022); Anderson v. Stability., No. 3:23-CV-00201 (N.D. Cal. filed Jan. 13, 2023); Getty Images, Inc. v. Stability AI, Inc., No. 1:23-CV-00135-UNA (D. Del. filed Feb. 3, 2023).

comparing and contrasting machine learning with the technologies examined in previous cases, finding that the case law strongly supports the conclusion that building and using generative ML models is either outside the scope of copyright or is a fair use.

1. A Primer on Machine Learning

One of the features of the US legal system is that the law is never analyzed in a vacuum. Legal opinions start by discussing the relevant facts of a case. Based on these facts, legal principles from previous cases are applied using logic and analogy, extending the law to new circumstances.[5]

Applying copyright law to machine learning should follow the same process. Unfortunately, most legal analyses in this area are incomplete or inaccurate in their descriptions of how ML models are made and used. These unsteady factual foundations have resulted in incorrect analogies and analyses in both lawsuits and law review articles.[6]

This section aims to provide an accurate and easily understandable description of the mechanics of machine learning. This foundation will then be used in later sections to analyze and apply relevant legal principles.

---

[5] Roscoe Pound, *Hierarchy of Sources and Forms in Different Systems of Law*, 7 TUL. L. REV. 475, 482-87 (1933) ("Principles do not attach any definite detailed legal results to any definite, detailed states of fact. . . . [They] are authoritative starting points for legal reasoning, employed continually and legitimately where cases are not covered or are not fully . . . covered by rules in the narrower sense.").

[6] Getty Images, Inc. v. Stability AI, Inc., No. 1:23-CV-00135-UNA, at *12 (D. Del. filed Feb. 3, 2023), https://fingfx.thomsonreuters.com/gfx/legaldocs/byvrlkmwnve/GETTY%20IMAGES%20AI%20LAWSUIT%20complaint.pdf) ("Stability AI encodes the images, which involves creating smaller versions of the images that take up less memory. Separately, Stability AI also encodes the paired text.").

Building and Training Machine Learning Models

The creation of a generative machine learning tool involves two phases. The first phase is the training of a "model." The second phase is using the "model" to make new outputs, such as new sentences, pictures, or code. These phases need to be examined separately because training and generation happen at different times, usually by different parties, and they involve different outputs.

Two Analogies for ML Training

One persistent misunderstanding some people have is how ML applications can recreate familiar objects. These people think of a machine learning model as just a complicated type of storage that saves everything it sees and then brings forth bits and pieces of memorized material to mash together into a collage.[7] In contrast, the power of machine learning is that it helps the computer identify meaningful correlations that are too attenuated or esoteric to be expressed by software developers. In other words, the model isn't memorizing the copyrightable expression in an input. Rather, it is evaluating and recording factual relationships between different elements of the expression.

Before diving into the mechanics of ML training, there are two analogies that may be helpful in developing a mental model of how ML training works: the art inspector and the law student. Both of these analogies illustrate the mechanics of model training, but in slightly different ways.

The Art Inspector

[7] *See, e.g.*, Anderson v. Stability., No. 3:23-CV-00201, at *3 (N.D. Cal. filed Jan. 13, 2023) (describing Stable Diffusion as "merely a complex collage tool.").

Imagine a newly hired art inspector given the job of examining every painting in the Louvre. This inspector has no background or experience in art and so has no preconceived ideas about art (what's beautiful or repugnant) or what is significant about any particular painting (what makes a Picasso a Picasso).

Lacking any guidance, the inspector studies each painting by measuring everything about it, such as the number of brushstrokes, paint thickness, average space between brushstrokes, size of the painting, and the thickness of lines. He includes *every* piece of information he can–the age of the painter, the date the painting was made, and in which corner the artist signed their name. The inspector measures aspects of the paintings that seem bizarrely random or unimportant, such as the number of consonants in the artist's name and the relationship between colors that are six inches apart. He is meticulous in his approach. Nothing is left untouched in the exhaustive analysis. Everything is recorded in the inspector's database.

As the inspector studies each painting, he tries to make his job more interesting by turning each measurement into a guessing game. Before he makes each measurement, he tries to predict what the answer will be, using the information he has gathered already. "How many brushstrokes are in this painting?" he wonders. "Well, it's a Rembrandt from the middle third of his career. I'd guess… 84 brushstrokes per square inch." The inspector then checks the measurement and records how good his prediction was before moving on to the next measurement and the next prediction. When the inspector begins to play, his answers are usually wrong. But as he takes more and more measurements, his predictions are increasingly correct.

After studying thousands (or millions) of paintings, the inspector is the world's foremost authority on validating paintings. He is regularly asked his opinion as to whether various newly-discovered paintings are legitimate. His ability to pinpoint which artist created a painting and to predict other things about each painting is unparalleled. Where before, the inspector took the artist's name and information to predict the measurements of their paintings, now the inspector uses the measurements to predict the painter.

The Law Student

When a student begins law school, they are frequently told that their job isn't just to learn the law–their job is to learn how to "think like a lawyer." As a result, legal teaching is structured differently than many other types of professional training. Learning the rules isn't enough; they must learn how to apply the law to new situations.

One common way of teaching legal reasoning is the *case method*. The case method involves studying judicial decisions, or cases, in order to understand the legal principles and rules that govern a particular area of law. Rather than simply memorizing legal rules and statutes, students learn to analyze and apply the law through a close examination of real-life legal disputes.

In practice, law students are given a set of facts, usually from a court decision, and then asked to analyze the legal issues raised. They look at the relevant laws, the arguments made by the parties, and the reasoning behind the decision. By examining the evidence and arguments, students develop their own understanding of the legal principles at play.

Law professors usually pair the case method with the Socratic method. In the Socratic method, the professor asks questions instead of providing answers. As the law

students struggle to imitate previous "correct" answers to similar questions, they begin to derive legal principles from the various scenarios. The professor provides feedback–validation of a correct analysis, or correction of a wrong answer–which the students then use to further refine their understanding.

When the time comes for the exam, a successful law student is able to take a hypothetical situation and generate a new analysis that nevertheless incorporates the correct principles, even though the student was never *explicitly* taught which principles to use. The student may not have a specific reason to weigh one factor over another, emphasize certain facts, or avoid certain arguments. She just knows, based on evaluation of example cases, how courts have weighted various facts and principles in the past.

In contrast, imagine a second student who attempts to master the material by memorizing all the facts and holdings from every case discussed in class. This second student does well when asked to describe the facts and holding of an important case, but fails to apply the principles of the law to new situations.

In short, the successful law student has a mental model of how the law is "supposed" to work based upon her analysis of the many cases studied during the class. Unlike the second student who just memorized facts, she can predict how courts would analyze new facts and new situations. She has learned to "think like a lawyer."

Examining the Two Analogies

Training an ML model is similar to the processes of the art inspector and the law student. In both cases, the basic steps are the same: *receive* an example; *predict* the relationship between the different elements of the example; *check* the result, and *adjust* to

improve future predictions. Those commonalities apply to the mechanical process performed by a computer during model training. However, there are some individual points from each analogy worth emphasizing.

The art inspector analogy is better at showing how ML models start with a clean slate. Before models are trained they are literally filled with random data. All associations that come out of ML applications were learned by observation. The art inspector's measurement of small, random details is also closer to the fixed process that occurs during ML training. But though the art inspector recorded and saved of all his measurements (inputs), what is actually recorded during ML training is instead the changing probabilities associated with different inputs.

The law student analogy is better at showing how unifying principles are the product of inference. Just like the law student is never explicitly taught the correct legal principles, ML training processes are not instructed what any of their inputs "mean." Instead, the "meaning" that is observed in an ML application is actually a complex probability function with millions or billions of parameters.

The law student analogy also demonstrates how direct memorization of inputs is actually antithetical to the goals of model training. Avoiding direct memorization is so important that ML training processes almost universally involve removing part of the training information to force the model to engage in the inference process. This is sometimes referred to as "masking" or "dropout." Failure to hide or remove information during training makes models unusable.

The danger in the law student analogy is that it makes the ML application easier to anthropomorphize. The process of training is not creative or selective. The ML models do not "think" or analogize. All the "learning" that occurs within the ML application is simply the rote construction or use of massively complex probability functions.

The Training Process

With these analogies in mind, we can examine how the same steps–receive, guess, check, adjust–are used to train an ML model. However, while the steps are conceptually similar to the mental processes of the art inspector and the law student, they are turned into a mechanical process that can be repeatedly performed by a computer.

Creating a "Brain": the Architecture of Machine Learning Models

Because computers do not have brains and senses like humans, the first step is to create a logical "brain"–a structure that can receive and process input. This structure is sometimes referred to as the "architecture" of the ML application.

To build a model, a data scientist begins by defining a logical structure for processing inputs to create outputs. Each part of the training process corresponds to a different part of the structure. These structures–initially inspired by the interconnections between brain cells–are called "neural networks." There are many different types of neural networks, but they share three general structures: an input layer, one or more "hidden" layers, and an output layer. These layers are made up of "nodes"–logical structures where values are temporarily stored and computation can occur. These nodes are highly interconnected by logical paths to other nodes. A stylized illustration is shown in Fig. 1.

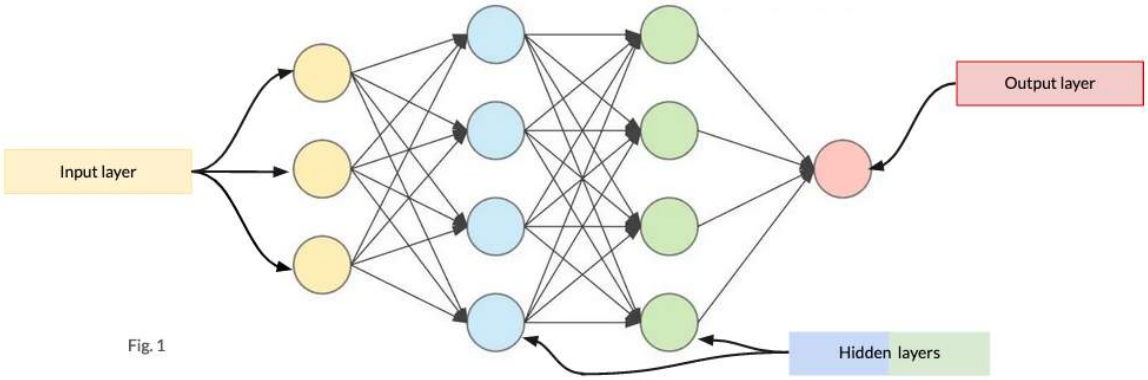


Fig. 1

This stylized figure has three nodes in the input layer, eight nodes in two hidden layers, and one node in the output layer. Different ML applications can have different numbers of nodes in each layer and can have many different types of interconnections.

The Input Layer (Receive)

The input layer of a neural network is where the data is provided to the model. It is similar to the art inspector viewing a painting or the law student reading a case.

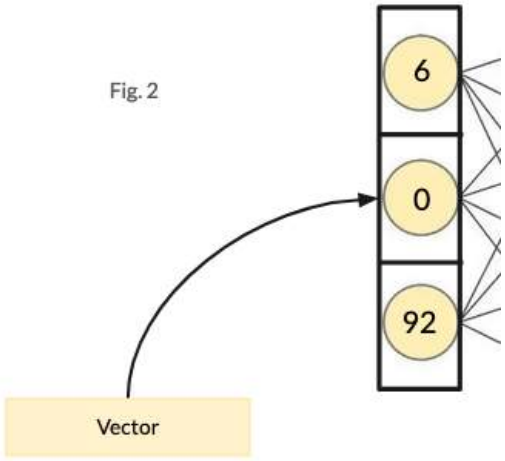
Unlike humans, who can process whole pictures or cases at a time, computers are more limited. Each node in the input layer has a memory designed to receive a single element of the input data. The goal of the input layer is to provide a uniform representation of the raw data that the model will use to make predictions.

As humans, we might think about these inputs as representing the pixels in an image or words on a page. However, from the computer's point of view, the input is just a list of numerical values called a "vector." For example, in a model that processes images, each node in the input layer just gets a number. Depending on the application, the number could represent the brightness of one part of an individual pixel. In a model that processes text,

each node in the input layer might receive a value that represents a word or character. See Fig. 2.

The Hidden Layers (Predict)

The hidden layers in a neural network are where the majority of the processing occurs in an ML application. These layers are called "hidden" because the data that is processed within them is not directly observable from the inputs or outputs of the model. The hidden layers contain a series of interconnected nodes, each of which performs a mathematical calculation on the inputs received from the previous layer. After performing the calculation, the node can pass forward the same value, a changed value, or nothing at all.



Each hidden node has an associated "weight" that changes the probability that a value will be changed or passed on. The weight corresponds to the model's "best guess" as to how the inputs should be used. See Fig. 3.

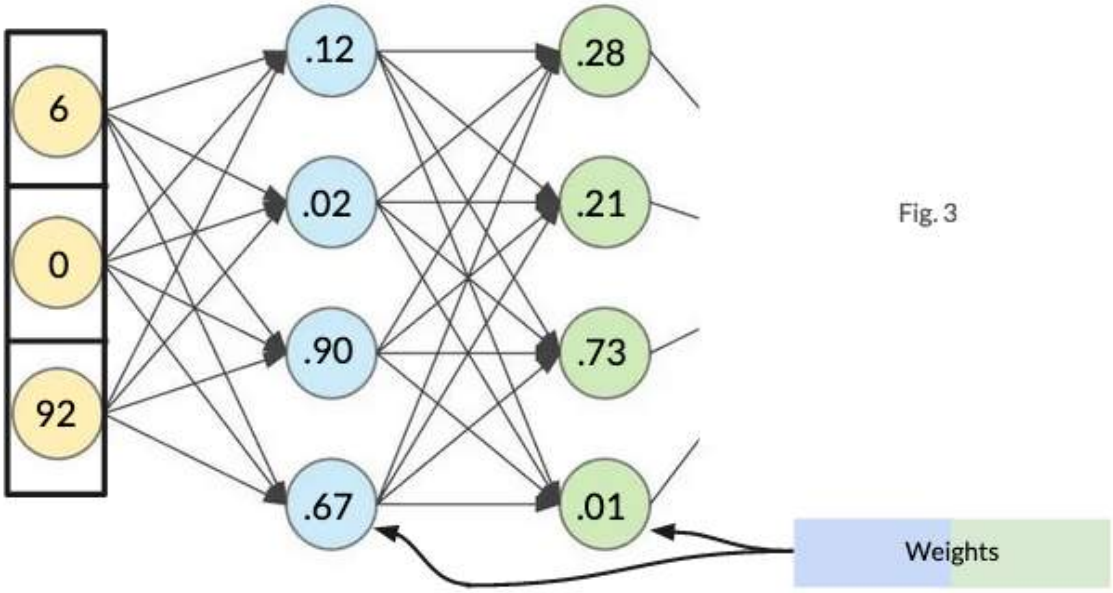


Fig. 3

Similar to the art inspector, scientists have no idea what information will end up being important for the evolution of the neural network. In the past, data scientists tried to identify specific "features" of the input data that they would provide to the neural network. However, isolating the right features took time, was error-prone, and didn't work as well. The current trend is simply to provide *all* of the data to the neural network and let the computer identify which correlations are useful. As a result, the correlations developed during the training process can be unexpected.

For example, in one application called "neural style transfer," some layers have been found to correspond to elements of an artist's style (like the number of brushstrokes, heaviness of lines, use of color) and other layers have been found to correspond to the large shapes and patterns in the image–what we would think of as the "content." The neural style transfer application generates a new output image by taking the "style" layers from a first

picture and the "content" layers from a second picture and using them together. See Fig. 4.

Fig. 4

Despite the identification of these correlations, neither the "style" nor the content of an image is saved as part of the model during training. Like the law student that extracts principles taught in court cases, the model has extracted correlations that, to humans, resemble certain artistic styles.

Applying this more concretely to the functioning of a real ML application, a common use of neural style transfer applications is to render pictures as if they had the unique brushwork of Vincent van Gogh. However, to perform this function, it would be counterproductive to save the brushstroke pattern for any existing van Gogh painting–none of the brushstrokes would fit a different image. Instead, it appears that one or more layers of the neural style transfer model contain a function that spreads out, moves, or changes the input values associated with each pixel in a way that creates a result that, to humans, resembles the brushwork style of van Gogh.

For humans might describe the process of creating a painting with van Gogh-style brushwork as "applying each color in a thick, contoured slab of paint." But for a model, this instruction might say like "for every pixel of blue, adjust the blue values of all surrounding pixels by an amount corresponding to this equation." It involves none of the

creativity and judgment that a human artist would need to use to apply the technique. Instead, it is "just" an evolved probability function that is literally inhuman in its complexity. Stable Diffusion, a modern model for processing and generating images, has 890 million parameters. GPT3, a model for understanding natural language, has 175 billion. Each parameter can be thought of as a conditional probability associated with one possible state of the neural network.

The Output Layer (Check)

The output of the hidden layers are then passed to the output layer, where the final result is provided. Just as with the input layers, the ML application doesn't "know" what the output represents. Like the input, it is just a vector of numbers. See Fig. 5.

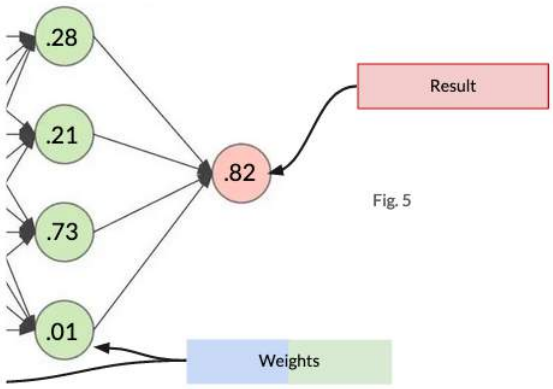


Fig. 5

How the result is interpreted is up to the human user. For example, the example value .82 in Fig. 5 could be interpreted as a classification ("there is an 82% chance this email is spam"), a recommendation ("if you liked this show, there is an 82% chance you might also like…"), or in the case of generative AI, one part of a pixel value of a new picture (change the blue in this output pixel to 82% of its maximum value) or the next word of a new sentence (the number .82 corresponds to the word "elephant").

Update the Weights (Adjust)

During training, every input has a known correct output (or possible output, if there are multiple correct possibilities). The result is compared with the correct output and the weights in the neural network are adjusted a tiny bit so that the next time the model receives a similar input, it will be more likely to provide a similar answer.

Try Again (Repeat)

This same process is then repeated. After processing and recording the predicted probabilities over the millions of provided examples, the application builds a comprehensive statistical picture of the range of possible answers for any given input and the likelihood of each answer. In many cases, the same inputs are re-used in different rounds of training to see if there are any further statistical correlations that can be learned from each example.

A portion (10-20%) of the input examples are never used as training inputs but are instead saved as a "testing" set. The testing set is never used as part of the training. The application receives the training input, predicts the output, and checks it against a known correct answer. However, the differences between the prediction and the correct answer are never used to adjust the weights in the model. Instead, the performance of the model is evaluated by using these never-used inputs as a barometer for how good the model's predictions have become.

At some point, the model's predictions stop improving. At that point, training is complete. The only way to further improve the model's prediction is to provide more example inputs for training. As more examples are used in the training process, the

resulting model has a better, more coherent picture of the interrelated probabilities required to process the inputs "correctly."

Defining the "Model"

In computer science terms, the interconnected network of weighted nodes is equivalent to a program–a very complicated program. **Training is the process of iteratively "evolving" the neural network's probabilities so it can emulate running a computer program that produces the correct outputs for each input.**

The "model," respectively, is the combination of the neural network design and the weights. **It is a set of numbers and equations that encode statistical probabilities about the inputs that have been processed during training.**

If a person were to save a well-trained model to disk and examine it, what that person would see would be a gigantic matrix of numbers–the learned weights associated with the nodes in the hidden layers. The model would not directly contain any copies of its source inputs, even in compressed form.

It is hard to say exactly what any single probability within the model represents. However, they provide a very detailed statistical picture of the collective experience of humanity when it comes to the inputs. In other words, the model doesn't really record knowledge about any single input. Instead, it records knowledge about what makes an image a *picture* as opposed to a bunch of noise, or what makes a bunch of words a *story* or an *article* instead of a bunch of gibberish. Or more specifically, what makes a song a "pop song from the 80s" as opposed to an "aria." And because humanity has embedded so much

implicit knowledge in words, pictures, and songs, these models can appear quite "intelligent."

The knowledge embedded within the model is "latent", in the sense that it is hidden, but unexpressed knowledge.. The logical landscape of this hidden knowledge–what concepts are "close together" or "far apart," and in which ways, is sometimes called the "latent domain."

Overtraining and Memorization

There is one specific type of training failure that is significant for copyright purposes: overtraining. If training continues beyond the level where the guesses stop improving, the probabilities associated with a specific input can get pinned to a specific output. The model still does not directly include its source inputs, but it has effectively "memorized" instructions for re-creating one or more inputs in response to a particular prompt. So while the model would not contain copies of the training works, it could nevertheless reproduce them if provoked to do so.

Overtraining and memorization is not the desired output of an ML model–it is a type of failure that scientists work to avoid. The desired outcome is a model that has encoded enough probabilities, that like the successful law student, it can respond effectively to novel inputs. An overtrained model is like the second law student that memorized the assigned cases but never learned to generalize. In other words, reproducing works is neither necessary nor desirable in machine learning.

In practice, overtraining in commercial models ranges from uncommon to extremely rare. For example, GitHub estimates that its CoPilot model for generating source

code includes a copied snippet longer than about 150 characters approximately 1% of the time.[8] Researchers studying the model for the Stable Diffusion image generator were able to make the model reproduce copies of about one hundred source images, only 0.0003% of the input training set.[9] And even then, those reproductions were hardly accessible. To find those reproductions, researchers concentrated on images that were duplicated hundreds of times in the dataset and then reconstructed the exact known parameters used to train the model for those duplicated images. Even with this head start, they still had to generate hundreds of *possible* duplicates and then use a specialized process to find the reported matches.

Inference and Generation Using Machine Learning Models

Once an ML model is trained, it can be applied to a task. Using an ML model is almost exactly the same as training an ML model. The difference is that there is no "Adjust" phase of the process. In use, the ML application *receives* the input and makes the same types of measurements as it would for any other input. It then uses the hidden layers with their trained weights to *predict* the output. However, once the prediction has been provided

[8] *Does GitHub CoPilot copy code from the training set?,* COPILOT: FEATURES (last visited Mar. 15, 2023), https://github.com/features/copilot ("Our latest internal research shows that about 1% of the time, a suggestion may contain some code snippets longer than ~150 characters that matches the training set. Previous research showed that many of these cases happen when GitHub Copilot is unable to glean sufficient context from the code you are writing, or when there is a common, perhaps even universal, solution to the problem.").

[9] NICHOLAS CARLINI, ET AL., EXTRACTING TRAINING DATA FROM DIFFUSION MODELS (2023), available at https://arxiv.org/abs/2301.13188. On page 6 the authors state that they studied the 350,000 most-duplicated images in the dataset and identified a total of 109 duplicated images, or 0.00031 percent. Taken as a percentage of the 5.85 billion images in the entire dataset, the identified percentage of duplicates found is 0.0000000018%.

to the output layer, the process is finished. The prediction is simply interpreted by the application and then returned as the result.

In general, there are two ways in which these ML results are used: *inference* and *generation*.

Inference

Inference refers to the process of using a trained model to make predictions or decisions based on new, previously unseen data. An ML application used for inference can usually be thought of as categorizing the newly-seen input in some way. This categorization can be interpreted as a classification, a prediction, or a recommendation. Thinking back to the art inspector, the inspector's ability to identify the artist associated with paintings he had never seen before is an example of inference. Because inference does not create a new work, it does not implicate copyright law.

Generation

Generation, on the other hand, refers to the process of creating *new* content, data, or outputs. Generation often involves models designed for tasks like natural language processing, image synthesis, or music composition.

Generative ML applications are usually designed to produce outputs of the same type as the inputs. For instance, a generative model trained on text data may be used to generate new text, such as sentences, paragraphs, or even entire articles. A model trained on images may be used to create new images. However, there is no inherent restriction forcing ML applications to generate outputs of the same type as their inputs. For example, some generative systems can take an image as an input and return a textual description of

the input, whereas other generative systems can take a textual description of a scene and return an image providing a rendering of the scene described by the user.

Just like inference, the generation process relies on the statistical patterns learned from the training data to create a predicted output. This output is returned as the result (or as part of the result). This is analogous to the law student's ability to create a new, coherent legal analysis based on the principles and lessons she derived from her case studies.

The difference between the law student and the ML application, however, is that the law student uses her intelligence and creativity to generate her answers, whereas an ML application has neither intelligence nor creativity. What the ML application does have is *context* and *randomness*.

Context

Taking the example of text, scientists have known since the 1960s that it was possible to construct sentences by analyzing a bunch of writing, finding which words tend to follow each other, and then repeatedly picking out the next word with the highest probability.

Humans instinctively perform this kind of analysis. For example, if someone was asked to predict the next word in the sentence "It was a dark and stormy _____," almost everyone would respond with the word "night." Sometimes there are a number of possible "next words," such as in the sentence "The wizard raised his _____." Some people might predict the next word might be "wand," "staff," or "hand."

When scientists tried to get computers to imitate humans, however, they quickly figured out that just choosing the most probable next word resulted in sentences that were

trite, ungrammatical, and repetitive. The difference was that the computer was only considering the single preceding word. Humans take into account all the words in the sentence, as well as the millions of words of context accumulated throughout our lives.

The logical way to improve the quality of the sentences created was to use more context when determining the most probable next word. Instead of only looking at the immediately preceding word, the computer could look at the two preceding words, the three immediately preceding words, or more. Nevertheless, using more than about five words of context usually resulted in systems that were too big to run on the computers of the time.

In the past fifteen years, however, the storage and processing capabilities of computers and networked computer systems have grown exponentially. As of the writing of this article, state-of-the-art text generation systems are able to take in about fifty typewritten pages of context when determining what word to generate next.[10] Scientists have also identified methods (called "attention") of helping the model adaptively use different parts of its provided context to improve generation.

Randomness

The second ingredient in generation is randomness. Scientists have discovered that one ingredient that makes humans creative is the element of surprise. Humans don't always

[10] GPT-4 Technical Report, https://arxiv.org/abs/2303.08774, What is the difference between the GPT-4 models? https://help.openai.com/en/articles/7127966-what-is-the-difference-between-the-gpt-4-models, OpenAI's GPT-4 is a safer and more useful ChatGPT that understands images, https://the-decoder.com/open-ai-gpt-4-announcement/ ("The context length of GPT-4 is limited to about 8,000 tokens, or about 25,000 words. There is also a version that can handle up to 32,000 tokens, or about 50 pages….").

use the highest probability outputs. We vary how we express ourselves in order to produce different effects on readers or viewers.

To emulate this tendency in humans, data scientists building generative ML applications include a parameter (frequently called "temperature") that is interpreted as a probability that the model should choose a slightly lower-probability path for a part of its output. For example, a temperature of 0.7 could mean that there is a 70% chance that the highest probability path will be used when generating an output, and there is a 30% chance that one of the lower probability paths will be taken instead.

The "temperature" used in an application does not correspond to any physical or logical law. It is a heuristic, derived over time and observation, that causes ML applications to seem more "human" in their outputs. Many ML applications allow users to control the temperature used for a particular generation. This allows a human using the ML application the ability to guide the course of generation by using or constraining the level of randomness affecting the output.

Controlling the Generation

Despite the use of limited randomness as part of the generative process, the output of an ML model is not random. A human using the ML application typically describes what should be generated and/or provides other inputs that are used to initialize and guide the generative process. These inputs are usually referred to as the "prompt."

The ML application takes the prompt and analyzes it as if it were an input. It then uses the analyzed prompt to identify a place in the latent domain to focus on when running the generative process. That is why when a user provides "cute purple dinosaur" into an

image generator, the application returns images of a cute purple dinosaur, not a motorcycle or a cloud. Further, the more information that is given within the prompt, the more control is exerted over the output.

The practice of developing a prompt that will give the desired output is sometimes referred to as "prompt engineering." Prompt engineering is actually an exploration through the latent space of the model–the probabilistic landscape of ideas and meanings–to match the generated expression to the author's or artist's conception. The goal of the author is to develop the exact set of inputs–images, words, and options–that will lead to the generation of the desired output.

2. Building Machine Learning Models Under US Copyright Law

There is no question of the importance of training material, including copyrighted material, for building cutting-edge ML systems. Using more training material results in better models, and better models mean better outputs. The leading ML models available today leverage billions of individual training examples, almost all of which are copyrighted. And yet, there is nothing improper with this usage. A comparison of cases and authorities with the actual mechanics of ML training suggests that in most cases, inputting copyrighted works into an ML model is a fair use, if it implicates copyright at all.

The Hypothetical

This is best observed by analyzing the training and distributing process of an actual ML model. This section will evaluate arguments for copyright infringement using the example of a "Stable Diffusion" model, as used in an ML-based image generation service,

to show that all uses of copyrighted material are either outside copyright's scope, de minimis, or covered by the "fair use" doctrine.

Stable Diffusion is a generative deep learning model that was released in 2022. It is designed to convert text descriptions provided by a user into images that match the artist's intent. It was trained using 5 billion images with matching text downloaded from websites on the Internet. Many of these images are commercially licensed.

Although this hypothetical uses real facts and is similar to real cases, it doesn't directly correspond to any one particular case. More details about those specific cases are available on the Internet. The applicable facts for this article are these: The training for the ML model was performed by a German university with funding from a for-profit UK-based company that has a United States affiliate. A copy of the model is provided by the German university to the U.S. entity. The U.S. company uses the model to generate images, for itself and for others. A copyright holder sues the U.S. entity for copyright infringement in the United States.

Threshold Questions–Does Copyright Even Apply to the Training Process?

Before evaluating whether building the model is a fair use, it's necessary to consider whether copyright even applies. The two specific issues for this question are 1) whether the training and import of the model is importation of material legally generated abroad, and 2) whether the copying of the image into the input during the training process is subject to limitations under 17 U.S.C. 117.

Importation of Information Legally Generated Abroad

One feature of the international race to create large ML models is that much of the training takes place in jurisdictions like India, the United Kingdom, and Germany. This is no accident. Unlike the United States, these countries (and others) have statutorily declared that using material for the purposes of ML training is not covered by copyright. Specifically, both the UK[11] and Germany[12] allow the reproduction of copyrighted works for non-commercial Text and Data Mining (TDM). As stated in the German Act:[13]

§ 60d on Text and Data Mining:

> For text and data mining on vast numbers of works (as source material) for scientific research, it is permitted:
>
> to reproduce the source material in an automated and systematic manner in order to create a corpus, especially through normalization, structuring and categorization;
>
> to make the corpus accessible to a defined circle of persons for joint scientific research as well as to individual third parties to check the quality of scientific research;

---

[11] Copyrights, Designs and Patents Act, (1988) § 29A, (1) (UK)

> (1) The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that—
>
> (a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and
>
> (b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).

[12] Urheberrechts-Wissensgesellschafts-Gesetz [Law on Copyright and Related Rights], Sep. 7, 2017, RGBl I at 3346, (Ger.), available at https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/BGBl-UrhWissG.pdf?__blob=publicationFile&v=1; https://www.clarin.eu/content/clic-copyright-exceptions-germany (English translation).

[13] *Id.*

> for non-commercial purposes only;
>
> the corpus and the copies of the source material shall be deleted after completion of the research and access to the public is to be terminated; however, it is permissible to transmit the corpus and reproductions of the original material to libraries and archives for permanent storage;

as far as the sui generis database right is concerned, such data mining as specified above is deemed lawful and therefore cannot be prohibited by the rightsholder.

The German TDM exception includes a number of procedural requirements, all of which must be met. Assuming the procedural requirements are met, the core intent of the law matches the behavior of the German university: "For… data mining on vast numbers of works, it is permitted… to reproduce the source material in an automated and systematic manner to create a corpus." An evaluation of the German TDM rule by the EU's Policy Department for Citizens' Rights and Constitutional Affairs concluded:[14]

The exception covers the acts of reproduction necessary for undertaking TDM…. It is worth noting that German law does not impose a "lawfully accessed source" requirement as France does. Also, it does not limit the source materials that can be mined to "text and data included or associated with scientific writings". With regard to databases, their reproduction is being qualified as constituting "normal use".[15]

---

[14] *THE EXCEPTION FOR TEXT AND DATA MINING (TDM) IN THE PROPOSED DIRECTIVE ON COPYRIGHT IN THE DIGITAL SINGLE MARKET - LEGAL ASPECTS* (Feb. 2018), https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA(2018)604941_EN.pdf.

[15] *Id.* at 18. The German TDM exception also allows for the creation and limited distribution of a "corpus," (e.g., source materials that were normalized, structured and categorized). However, as the model creation process does not create a corpus, this provision is inapplicable.

If the creation of the model is *per se* lawful in Germany, then importation of the model into the US might be allowed under *Kirtsaeng v. John Wiley & Sons, Inc*., 568 U.S. 519.

Kirtsaeng involved a dispute between a student (Kirtsaeng) and an academic textbook publisher (Wiley). Wiley assigned the rights to publish, print, and sell foreign editions of its English language textbooks abroad to its wholly-owned subsidiary, Wiley Asia. The books were sold at low prices in Thailand,where Kirtsaeng bought them, and then shipped to the United States where he re-sold them for a profit.

Wiley filed a lawsuit against Kirtsaeng, claiming that his unauthorized importation and resale of its books was an infringement of its exclusive right to distribute and prohibit importation. However, Kirtsaeng argued that the "first sale" doctrine permitted the importation and resale of the books as they were "lawfully made" and acquired legitimately.

The District Court and Second Circuit held that the "first sale" doctrine does not apply to goods manufactured abroad. The U.S. Supreme Court reversed, finding that any works "subject to protection under this title" included works "without regard to the nationality or domicile of the author," and that any works "first published" in any nation that had signed a copyright treaty with the U.S. could be considered "lawfully made under [the Copyright Act]."[16]

Kirtsaeng is usually cited relative to the first sale doctrine under copyright law. However, the court's holding is not limited to first sale only. The court said that the

[16] *Id.*

exclusive right of distribution under copyright "is by its terms "[s]ubject to" the various doctrines and principles contained in §§107 through 122"[17] which specifically apply to all works "lawfully made under this title" - which includes Berne signatories like Germany.[18]

The German university's actions under the German TDM exception meet this standard of "lawfulness."[19] The text of the exception emphasizes that the University's behavior "**is deemed lawful** and therefore cannot be prohibited by the rightsholder." Principles of comity suggest that US courts give heed to German copyright law when it is applicable, as here, to determine whether a particular copy was lawfully made.

In *Kirtsaeng*, the importation of the textbooks was lawful because the publisher's exclusive rights in those copies of the books had been exhausted. In contrast, the German TDM exception excludes the creation of the model from copyright altogether. The similarity is that in both *Kirtsaeng* and in this hypothetical, the use of the material "cannot be prohibited" by the rightsholder.[20]

Limitations on Exclusive Rights in Computer Programs

Putting aside the German TDM exception, a second hurdle is the possible application of 17 U.S.C. 117(a) to the copyrighted input material. 17 U.S.C. 117 of the United States Copyright law lays out exceptions to the exclusive rights of the copyright owner. The original text of this section confirmed the application of copyright to software, stating that the Act did not give the owner of a copyright in a software work "any greater

[17] Kirtsaeng v. John Wiley & Sons, Inc., 568 U.S. 519, 524, 133 S. Ct. 1351, 1355 (2013).
[18] *Id.*
[19] Kirtsaeng v. John Wiley & Sons, Inc., 568 U.S. 519, 530, 133 S. Ct. 1351, 1358 (2013).
[20] German TDM exception.

or lesser rights" when used in conjunction with a machine.[21] In 1980, this section was amended to better address computer technology, following the recommendations of the National Commission on the New Technological Uses of Copyrighted Works (CONTU).[22] One of the changes was the addition of section 117(a)(1). Section 117 states in relevant part:

> **(a) Making of Additional Copy or Adaptation by Owner of Copy.**— Notwithstanding the provisions of section 106, it is not an infringement for the owner of a copy of a computer program to make or authorize the making of another copy or adaptation of that computer program provided:
> (1) that such a new copy or adaptation is created as an essential step in the utilization of the computer program in conjunction with a machine and that it is used in no other manner….

Two immediate questions present themselves. First, whether a digital copy of a work, like the images used to train Stable Diffusion, is a "computer program," and 2) whether a person that downloads a copy of an image from a website is an "owner" of that copy.

Digital Works as Computer Programs

Many people naturally draw a distinction between a computer program and a computer file due to the common use of these terms. However, the definition in section 117 is not so limited. It defines a computer program as "a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result." It may seem surprising to some, but this definition includes digital media files.

[21] Pub. L. 94–553, title I, §101, 90 Stat. 2546 (1976) (codified at 17 U.S.C. §117).

[22] NAT'L COMM'N ON NEW TECH. USES OF COPYRIGHTED WORKS, PB85-225621, FINAL REPORT OF THE NATIONAL COMMISSION ON THE NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS (CONTU Rep.) (1978) at 12.

When we think about an image or a song on our computer, we usually don't pay attention to the distinction between how a file is encoded and the copyrighted work itself. However, the files we use on our computer are not the work, just as a compact disk is not the same as the music *on* the compact disk.

Turning specifically to images, when a person views an image on a computer, the image they are viewing consists of individually controlled pixels on the computer screen that create the pattern of light and color we recognize as a picture. This is the "certain result" envisioned by the statute. We are likely familiar with many of the common file extensions, such as .jpg and .png. Two image files with different extensions may result in identical images being placed on the screen, but the process followed by the computer is different for every type of file.

Each different type of image file contains a different type of instructions that are interpreted by the computer processor to understand how to control the pixels in the screen to create the desired output. These are the "set of statements or instructions to be used directly or indirectly in a computer" to bring about the result (the showing of the picture).

These instructions are usually converted to a binary format for ease of use on a computer–but not always. One type of image format (.ps, for "Postscript") consists only of human-readable instructions that tell the computer how to draw the desired image on the screen.

Because image files have "a set of statements or instructions" that is used by a computer to bring about "a certain result," image files–like all digital media–fit the statutory definition of a "computer program."

Ownership of a "Copy" of a Work

One limitation of 117(a) is that its application is limited to "the owner of a copy of a computer program." Most standard computer programs, like Microsoft Word or Mozilla Firefox, are licensed, not sold. Each download or installation is preceded by an "End User License Agreement" screen laying out the terms of use under the copyright owner's license.

However, the situation is different for the content available on websites. There is no license agreement for each file retrieved from a website and no opportunity for meaningful offer and acceptance. By design, people put content on the Internet so that users and automated processes can receive copies of the website contents and view them on their computer. As explained, this process requires the computer to implement the file "instructions" and thereby create an individual copy of the work.

The reproduction of these copyrighted works is by design. Copyright owners have the option to not put their content on the Internet or to impose controls like password-protecting certain files. Copyright owners forgo these protections knowing that, absent controls that limit access, Internet websites are designed to automatically provide copies of all requested files. Websites provide these files even before humans are able to perceive

the results or agree to any purported terms of use. Copyright owners *intend* this transfer to occur.[23] Without it, no one would be able to view their website.[24]

Therefore, the most common occurrence when placing content on the Internet is that a copy is automatically provided to any person or agent wanting to receive a copy. This copy is provided by the copyright owner or an authorized licensee for use on a computer to read or view the copy. This transfer may not involve a monetary payment, but website owners usually expect to benefit in other ways. This expected benefit may take the form of advertising revenue, brand improvement, future sales, or even just increased recognition.[25] Under common-law principles, a good given for free with the expectation of

[23] Kernal Records Oy v. Mosley, 794 F. Supp. 2d 1355, 1364 (S.D. Fla. 2011) (quoting Getaped.com, Inc. v. Cangemi, 188 F. Supp. 2d 398, 402 (S.D.N.Y. 2002)) ("Consequently, when a website goes live, the creator loses the ability to control either duplication or further distribution of his or her work. A webpage in this respect is indistinguishable from photographs, music files or software posted on the web - all can be freely copied. Thus, when a webpage goes live on the Internet, it is distributed and 'published.'"); *see also* Playboy Enters. v. Russ Hardenburgh, 982 F. Supp. 503, 513 (N.D. Ohio 1997) (plaintiffs violated the copyright holder's publishing right by moving password protected images onto a publicly accessible webpage). *Playboy* makes clear that if a copyright owner forgoes protecting their works by making them publicly available, then they've consented to individual copying. *Id.*

[24] There is a standard, called robots.txt, that requests that automated Internet agents refrain from downloading certain files or requesting the contents of particular URLs. Most automated Internet agents respect this standard, but compliance is optional. The content still remains available for download at any time. *See* Google Search Central, *Robots FAQs*, GOOGLE SEARCH CENTRAL (last updated Feb. 20, 2023), https://developers.google.com/search/docs/crawling-indexing/robots/robots-faq; Field v. Google, 412 F. Supp. 2d 1106, 1113 (D. Nev. 2006).

[25] See, e.g., True Freight Logistics LLC v. Glob. Tranz Enters., No. CV-18-01472-PHX-JGZ, 2019 U.S. Dist. LEXIS 237148, at *10 (D. Ariz. Jan. 9, 2019)(Getting traffic sent to a company's website was part of a "party's reasonably expected benefits of the bargain.").

benefit can still be a "sale."[26] Accordingly, the recipient of the copy is the owner of *that copy* of the copyrighted work.[27]

The recipient's ownership of this single copy of the work doesn't significantly undermine the exclusive rights of the copyright owner. The recipient doesn't own the *copyright* in the work. The recipient receives no rights to reproduce, sell, or distribute the work. The one exception are the rights granted under section 117, which include the right to make a copy if doing so is an essential step in the utilization of the work on a computer.

Under this analysis, the use of retrieved material for the limited purpose of ML model training is one of the few situations that fit squarely into the confines of section 117(a)(1). The individual who accesses copyrighted work online obtains a limited ownership in that work, which they can copy if it is essential in utilizing a work on a computer. Since ML models qualify as "computer programs," and inputting copyrighted works is necessary for them to function, then ML training on copyrighted works is outside the scope of the exclusive rights granted under copyright law.

---

[26] "Defendant was offering complimentary drinks to its patrons. Nonetheless, it was not offering these drinks out of any sense of hospitality or charity. Defendant runs a casino, and the complimentary drinks were offered as an incentive to patrons to gamble, and therefore enhance defendant's business.",Levondosky v. Marina Assocs., 731 F. Supp. 1210, 1212 (D.N.J. 1990).

[27] Hamer v. Sidway, 124 N.Y. 538, 546 (Ct. App. N.Y. 1891) ("Consideration means not so much that one party is profiting as that the other **abandons some legal right in the present or limits his legal freedom of action in the future as an inducement** for the promise of the first.") (emphasis added). In *Hamer* an uncle promised to give ownership of pecuniary property ($5,000) to his nephew if he would abstain from certain activities until he was twenty-one. Though the uncle gained no monetary or tangible benefit, the court held that it was a valid contract. *Id.* at 551. Corbin on Contracts supports this saying that "there are innumerable transactions, even including many that are called commercial, in which the promisor receives nothing of economic advantage, the promisor receives no "benefit" that is measurable with money or even with other things of value." 2 Corbin on Contracts § 5.9 (2022).

Capitol Records, LLC v. ReDigi Inc., 934 F. Supp. 2d 640 is not to the contrary. The *ReDigi* case dealt with resale of digital goods. In the circumstance described above, however, the transaction is between the copyright holder and an immediate recipient. *ReDigi* likely restricts the ability of the recipient to resell any of the content received over the Internet–but it does not prohibit first-party interaction.

The Application of Copyright to Building the Model

But what if a court were to incorrectly hold that ML training was covered by copyright law? Under that assumption, every use of the copyrighted material–even if it is shared with a machine rather than other persons–would presumptively violate the copyright owner's exclusive right to reproduce a work under section 106 of the Copyright Act.[28]

Separate, but related to the training issue, is whether the completed product (the "trained" model) would also violate copyright law through its outputs. If the model output closely resembles a copyrighted work in the training dataset, it could constitute infringement by either being a derivative work, if not a complete reproduction.[29] Though separate problems, the question for both of these issues is the same: whether these acts are excused as fair use under section 107.

The Fair Use Standard

The fair use doctrine is designed to balance the protection that copyright law grants to owners with the greater public good and to encourage creativity, education, and free speech.[30] This doctrine allows for the use of copyrighted materials without the permission of the owner for specific purposes such as criticism, comment, news reporting, teaching,

[28] 17 U.S.C. §106 ("the owner of copyright under this title has the exclusive rights to do and to authorize any of the following:(1) to reproduce the copyrighted work in copies or phonorecords; (2) to prepare derivative works based upon the copyrighted work….").

[29] *Id.*

[30] Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 95 (2d Cir. 2014) ("*there* are important limits to an author's rights to control original and derivative works. One such limit is the doctrine of "fair use," which allows the public to draw upon copyrighted materials without the permission of the copyright holder in certain circumstances.").

scholarship, or research.[31] The determination of whether a particular use is considered fair use is decided on a case-by-case basis and is a combination of law and fact.[32] There is no automatic assumption of fair use. Fair use is an affirmative defense, with the defendant bearing the burden of proof.[33]

Fair use is evaluated based on four factors: 1) the purpose and character of the use; 2) the nature of the copyrighted work; 3) the amount or substantiality of the portion used; and 4) the effect of the use on the potential market or value of the work.[34] Courts have also considered whether a particular use advances the public purpose of encouraging the creation of new works.[35]

The Purpose and Character of the Use

Regarding the first factor of fair use, the purpose and character of the use, several aspects of ML training are directly relevant to the inquiry. These are: 1) the use is transformative; 2) the use is limited to "reading" the work; 3) the work is only used for making measurements and recording facts about the content; and 4) the use is for a research purpose.

---

[31] See e.g., Folsom v. Marsh, 9 F. Cas. 342, 344, F. Cas. No. 4901 (C.C.D. Mass. 1841) (criticism and comment); Nunez v. Caribbean Int'l News Corp., 235 F.3d 18, 25 (1st Cir. 2000) (publishing copyrighted images for new reporting); Cambridge Univ. Press v. Patton, 768 F. 3d 1232, 1242 (11 Cir. 2014); Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 95 (2d Cir. 2014) (scholarship); Sony Comp. Entertainment, Inc. v. Connectix Corp., 203 F.3d 596, 599-601 (9th Cir. 2000) (using copyrighted software for reverse engineering, i.e. "research.").

[32] Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 560 (1985).

[33] Am. Geophysical Union v. Texaco Inc, 60 F. 3d 913, 918 (2d Cir. 1994).

[34] Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 96 (2d Cir. 2014).

[35] *Id.* at 94 (quoting Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 575 (1994)) ("for fair use of copyrighted materials has been thought necessary to fulfill copyright's very purpose, "to promote the Progress of Science and useful Arts…."); see also U.S. Const., Art. I, § 8, cl. 8. ("To 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.").

The Use in ML Training is Transformative

The primary consideration relevant to the character and use of the work is whether the use of the work is transformative. The concept of transformative use comes from the 1994 Supreme Court decision in *Campbell v. Acuff-Rose Music*.[36] In *Campbell*, the Supreme Court described "transformative" use as being the key element underlying the first fair use factor:

> The central purpose of this investigation is to see, in Justice Story's words, whether the new work merely "supersede[s] the objects'" of the original creation, or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message; it asks, in other words, whether and to what extent the new work is "transformative."[37]

The leading cases regarding computer-driven transformation of works are *Authors Guild, Inc. v. HathiTrust*[38] and the related case *Authors Guild* v. *Google*.[39] The factual background of these two cases is similar: the Authors Guild sued HathiTrust and Google for copyright infringement because of the defendants' mass digitization of books. Hathitrust, a digital library consortium, created its digital copies for preservation, for accessibility for visually impaired users, and to create a search index. Google created Google Book search to facilitate researchers in locating relevant information. Users could search across books for specific words and phrases and then see a "snippet view" with the search result highlighted.

[36] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 586, 114 S. Ct. 1164, 1175 (1994).
[37] *Id*. at 579.
[38] Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 97 (2d Cir. 2014).
[39] Authors Guild, Inc. v. Google, Inc., 804 F.3d 202, 217 (2d Cir. 2015).

In these cases, there were two accused processes: 1) the creation of a digital copy of the books for the purposes of creating a search index, and 2) the distribution of whole or partial copies to users. Since the process of training an ML model does not result in distribution, the relevant portion for our purposes is the creation of the digital copy for indexing.

In the *Authors Guild* cases, digitization was accomplished by "mak[ing] a digital scan of each book, extract[ing] a machine-readable text, and creat[ing] an index of the machine-readable text of each book."[40] The end result was a search index enabling users to find content within the books more effectively as well as research new types of questions.[41]

With regard to the creation of the search index, the *Hathitrust* court said:

> [We] conclude that the creation of a full-text searchable database is a quintessentially transformative use…. the result of a word search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn. Indeed, we can discern little or no resemblance between the original text and the results of the [defendant's] full-text search.[42]

The *Google* court further elaborated on the transformative nature of the search index by highlighting the new statistical research tools that it made possible:

> [The] purpose of Google's copying of the original copyrighted books is to make available significant information about those books, permitting a searcher to identify those that contain a word or term of interest, as well as those that do not include reference to it. In addition, through the ngrams tool, Google allows readers to learn the frequency of usage of selected words in the aggregate corpus of published books in different historical periods. We have no doubt that the purpose of this copying is the sort of

[40] Authors Guild v. Google, Inc., 804 F.3d 202, 208 (2d Cir. 2015).
[41] *Id.*
[42] Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 97 (2d Cir. 2014).

> transformative purpose described in *Campbell* as strongly favoring satisfaction of the first factor.[43]

The *Authors Guild* cases are not alone. Many other courts looking at similar fact patterns have found the same. For example, in A.V. v. iParadigms, LLC, the court found that the copying and archiving of student papers is permissible when aimed at detecting and preventing plagiarism rather than capturing expressive content. In Perfect 10 v. Amazon.com, Inc., the court ruled that Google's copying of Internet content to make it searchable was considered transformative as it turned the image into a "pointer" directing the user to a source of information. Similarly, in Kelly v. Arriba Soft Corp., the court ruled that copying to produce exact replicas of artistic works displayed in thumbnail form on the internet was transformative as it was unrelated to any aesthetic purpose and was aimed at facilitating searches.

Just like the building of a search index is a "quintessentially transformative use," so too is the building of an ML model. The result of the machine-based processing is a product with wholly different purposes, capabilities, and uses. There is no way in which an ML model could be mistaken for any of its training inputs. The mass of statistical probabilities that make up a generative ML model are so different from the training material that there is no question it is "different in purpose, character, expression, meaning, and message"[44] from any (or all) of the works that were used as input.

Also significant is that ML models, like the search index in the *Authors Guild* cases, records information *about* the works used for training, not any of the expression contained

[43] Authors Guild v. Google, Inc., 804 F.3d 202, 217 (2d Cir. 2015).
[44] Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 97 (2d Cir. 2014).

within the works themselves. And, just like the search index in *Google*, ML models allow people to perform new types of research and discover new correlations.

The Use in ML Training is Limited to "Reading" the Work

One persistent misunderstanding is the perception that the ML training process makes repeated or derivative copies of each work used as input.[45] There is also the perception that the model somehow "stores" the works used for training within the model. Both of these perceptions are incorrect.

As described relative to the "receive" part of ML training, there is only one copy of the work needed for training: the initial copying of the work into the input layer of the ML model. This is the process by which the model "reads" the input in order to perform the training process. Reading (or viewing) a work, including on or by a computer, has been repeatedly found to be fair use.

---

[45] *See* Benjamin L.W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM.. J.L. & ARTS 45, 48 (2017) (These "training data" often comprise thousands of unauthorized copies of copyrighted works, which are reduplicated and modified countless more times throughout the training process."); *id.* at 62 ("Once an input dataset has been compiled, it may be copied, emulated, and re-copied thousands of times during the learning process."); U.S. PAT. & TRADEMARK OFF., PUBLIC VIEWS ON ARTIFICIAL INTELLIGENCE AND INTELLECTUAL PROPERTY POLICY (2020) (saying that ML "functions by ingesting copyright works" which results in "mass digitization.").

Courts do not seem to have an incorrect notion of ML, rather they often have no notion at all. *See* Carpenter v. McDonald'S Corp., 580 F. Supp. 3d 512, 516 (N.D. Ill. 2022) (describing ML simply as "a form of artificial intelligence."); Performance Pricing, Inc. v. Google, Inc., No. 2:07cv432, 2009 U.S. Dist. LEXIS 77538, at *5 fn. 3 (E.D. Tex. Aug. 28, 2009) (describing ML as "a type of computational algorithm which is derived by other algorithms."). *But see* Ocean Tomo, LLC v. Patentratings, LLC, 375 F. Supp. 3d 915, 956 (N.D. Ill. 2019) ("At a high level, machine learning tools attempt to discern patterns within data, but with no pre-conceived concepts or requirements as to the structure of these data. Machine learning uses an iterative process, in which the system initially forecasts an outcome based on combinations of input variables. The system then determines the errors of its forecasts, and adjusts accordingly, iterating until these error terms are minimized.").

For most of history, the idea that receiving or reading a work might implicate a copyright owner's exclusive rights would have seemed absurd.[46] The exclusive rights granted to the copyright owner only address the means of reproduction and distribution. They do not include the right to read the work, which has been and still is unrestricted.[47] This is consistent with the overriding purpose of the patent and copyright clause of the U.S. Constitution: "to promote the Progress of Science and useful Arts."[48] The widespread dissemination of knowledge is the underlying policy purpose for all copyright law.[49]

It is instructive to compare the exclusive audiovisual rights granted to copyright owners under 17 U.S.C. 106(4)-(6). These subsections address the rights of public display, performance, and broadcast. Even though it is receiving the work that drives demand, the

---

[46] *See* Jessica Litman, The Exclusive Right to Read, 13 CARDOZO ARTS & ENT. L.J. 29, 34 (1994) ("Ninety years later, the U.S. copyright law is even more technical, inconsistent and difficult to understand; more importantly, it touches everyone and everything….Most of us can no longer spend even an hour without colliding with the copyright law. Reading one's mail or picking up one's telephone messages these days requires many of us to commit acts that the government's Information Infrastructure Task Force now tells us ought to be viewed as unauthorized reproductions or transmissions.")); Jessica Litman, *Readers' Copyright*, 58 J. COPYRIGHT SOC'Y U.S.A. 325 (2010) ("Copyright gives no exclusive rights to control private performance or display.80 What you do with a book, movie, or sound recording in your living room is not copyright infringement, even if your copy is pirated. Private performance and display is simply off limits. (That isn't because copyright owners didn't ask for private performance and display rights - they did. But nobody took those demands seriously, I think, because at some level everyone understood that the freedom to read and enjoy material without the copyright police looking over your shoulder is an interest that copyright law has respected and should protect.").

[47] Jessica Litman, *Lawful Personal Use*, 85 TEX. L. REV. 1871, 1882 (2007) ("copyright . . . left reading, listening, and viewing unconstrained.").

[48] U.S. CONST., Art. I, § 8, cl. 8.

[49] Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975) ("Creative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts…. 'The sole interest of the United States and the primary object in conferring the monopoly,' this Court has said, 'lie in the general benefits derived by the public from the labors of authors.'"); Cambridge Univ. Press v. Patton, 769 F.3d 1232, 1237 (11th Cir. 2014) ("These boundaries must be drawn carefully in order to assure that copyright law serves its intended purpose, which is to promote the creation of new works for the public good by providing authors and other creators with an economic incentive to create.").

exclusive rights granted under the law are focused on the making available of the work, not its reception.

It is only with the advent of computers that reading has been brought within the ambit of copyright owners' control due to the fact that at least one transient copy (sometimes referred to as a "RAM copy") is technically required for any user to perceive or use a work on a computer.

A few widely-criticized decisions, mostly in the Ninth Circuit, have found that the RAM copy is enough to support a charge of copyright infringement.[50] In contrast, most scholars and courts have found that there is a fair use right to receive and "read" a work, even if that reading necessarily involves creating a copy.[51]

---

[50] MAI Sys. Corp. v. Peak Comput., Inc., 991 F.2d 511, 519 (9th Cir. 1993) ("However, since we find that the copy created in the RAM can be 'perceived, reproduced, or otherwise communicated,' we hold that the loading of software into the RAM creates a copy under the Copyright Act.") ; MDY Indus., LLC v. Blizzard Entm't, Inc., 629 F.3d 928, 938 (9th Cir. 2010) ("The parties agree that when playing WoW, a player's computer creates a copy of the game's software in the computer's random access memory ("RAM"), a form of temporary memory used by computers to run software programs. This copy potentially infringes unless the player (1) is a licensee whose use of the software is within the scope of the license or (2) owns the copy of the software.") ; *see also* Vault Corp. v. Quaid Software Ltd., 847 F.2d 255, 260 (5th Cir. 1988) ("the act of loading a program from a medium of storage into a computer's memory creates a copy of the program."). As for criticism see 2 Nimmer on Copyright § 8.08 ("However, it is submitted above that MAI v. Peak itself wrongly concluded in favor of liability.".

[51] At one point courts considered RAM copies to be copyright infringement but they now view it as an "implied license" when "the copyright holder knows of the use and encourages it." Field v. Google, 412 F. Supp. 2d 1106, 1116 (D. Nev. 2006). The purpose of publishing content on the Internet is for people to view it, and no one can view it unless their computer makes a copy. Ticketmaster L.L.C. v. RMG Techs., Inc., 507 F. Supp. 2d 1096, 1105 (C.D. Cal. 2007) ("[C]opies of webpages [are] stored automatically in a computer's cache or random access memory ("RAM") upon a viewing of the webpage."). Copyright owners expect and want these copies to be made. *Field* 412 F. Supp. 2d at 1114. The leading case on this is Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984). *See also* Fox Broad. Co. v. Dish Network L.L.C., 747 F.3d 1060 (9th Cir. 2013); Recording Indus. Ass'n of Am. v. Diamond Multimedia Sys., Inc., 180 F.3d 1072 (9th Cir. 1999).

See James Grimmelmann, *Copyright for Literate Robots*, 101 IOWA L. REV. 657, 659 (2016) (quoting Jessica Litman, *Lawful Personal Use*, 85 TEX. L. REV. 1871, 1882 (2007)) ("In a world of books and other pre-digital technologies, 'copyright . . . left reading, listening, and viewing unconstrained.'"); Jessica Litman, Lawful Personal Use, 85 TEX. L. REV. 1871, 1897–903 (2007) (listing many examples of day-to-

Sony Corp. of Am. v. Universal City Studios, Inc.[52] is instructive. The case centered around Sony's Betamax video cassette recorder (VCR), which allowed users to record television programs for later viewing, a practice known as "time-shifting." Universal City Studios, along with other movie studios, sued Sony, arguing that the VCR facilitated copyright infringement by enabling users to record copyrighted television programs without authorization. The plaintiffs sought monetary damages and an injunction to stop the production and sale of Betamax VCRs.

The Supreme Court, in a 5-4 decision, ruled in favor of Sony, stating that noncommercial home use of the VCR to record television programs for later viewing is fair use. Although time-shifting required making a copy of a copyrighted work, it was non-infringing because the purpose was to allow the users to receive the work at the time of their choosing—not distribution, publishing or performance.[53]

This fair use "right to read" is illustrated by the use of a web browser to read online materials. Just by opening a webpage, people request copyrighted works from the owners

day private copying that violate the strict language of the copyright statute, yet are protected as fair use); Aaron Perzanowski & Jason Schultz, *Copyright Exhaustion and the Personal Use Dilemma,* 96 MINN. L. REV. 2067, 2086–2092 (2012) (defending personal uses as fair use, including the right to read); C. Edwin Baker, *First Amendment Limits on Copyright*, 55 VAND. L. REV. 891, 904 (2002) ("The expressive liberty protected by the First Amendment encompasses copying as a way of receiving or preserving personal access"); Jed Rubenfeld, *The Freedom of Imagination: Copyright's Constitutionality*, 112 YALE L.J. 1, 38 (2002) ("Because it protects the freedom of imagination, the First Amendment directly protects not only speakers, but readers, viewers, and *listeners* as well.") (emphasis added); Diane Leenheer Zimmerman, *Is There a Right to Have Something to Say? One View of the Public Domain*, 73 FORDHAM L. REV. 297, 326 (2004) ("Speech requires… some ability to acquire such content and certainly the privilege of using it.").

[52] Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984).

[53] *Id.* at 449 ("time-shifting merely enables a viewer to see such a work which he had been invited to witness in its entirety free of charge, the fact that the entire work is reproduced, does not have its ordinary effect of militating against a finding of fair use."). The defendants in Sony brought substantial evidence that the copyright holders wanted users to make copies, if necessary to view their works. *Id.* at 445 (Fred Rogers, copyright holder of Mister Rogers' neighborhood testified that "he had absolutely no objection to home taping for noncommercial use and expressed the opinion that it is a real service to families to be able to record children's programs and to show them at appropriate times.").

and make a RAM copy on their computer in order to perceive the work. This happens millions of times every day. It has never been questioned whether ordinary web browsing is fair use. Any attempt to sue end users for making an incidental, necessary copy of a work, freely provided by the owner in response to a web request, would be quickly and easily disposed of as non-infringing.

Courts have noted and emphasized that this type of copying is fair use. "[M]erely by accessing a webpage, an Internet user acquires the ability to make a copy of that webpage."[54] Like software, "copies of webpages [are] stored automatically in a computer's cache or random access memory ("RAM") upon a viewing of the webpage."[55] Yet this copying is non-infringing, because, there is an "implied license" to make copies when "the copyright holder knows of the use and encourages it."[56]

The use of copyrighted materials as input to an ML model is exactly the same as the use of copyrighted materials as input to a web browser. Both recipients receive a copy of the work and view it by loading it into memory so that it can be processed by the computer. The only difference is that in the web browser, it is a human doing the viewing and in ML training, it is the machine that "views" the data.[57]

ML Model Training is Limited to Making Measurements and Recording Facts

[54]Getaped.com, Inc. v. Cangemi, 188 F. Supp. 2d 398, 402 (S.D.N.Y. 2002).

[55] Costar Realty Info., Inc. v. Field, 737 F. Supp. 2d 496, 507 (S.D. Md. 2010).

[56]Field v. Google, 412 F. Supp. 2d at 1115 (D. Nev. 2006); *see also id.* at 1114 ("Field knew that if he used the "no-archive" meta-tag on the pages of his site, Google would not provide "Cached" links for the pages containing his works. Field consciously chose not to use the "no-archive" meta-tag on his Website….When the pages containing Field's copyright works were displayed in Google's search results, they were automatically displayed with "Cached" links, as Field intended they would be.").

[57] To the extent that reading for personal edification is different than a machine "reading" for the purpose of generating a statistical model, courts have almost uniformly found that machine "reading" is fair use. *See generally* James Grimmelmann, *Copyright for Literate Robots*, 101 Iowa L. Rev. 657, 659 (2016).

Also relevant to the first fair use factor is that ML training is limited to making measurements and recording facts. Because the outputs of ML applications *seem* so expressive, people mistakenly assume that ML applications copy creative expression from inputs and use the copied expression to generate derivative outputs.[58]

Rather than copying any expression, however, the model training process records facts about the work. Think of the analogy of the art inspector taking every measurement possible–brushstrokes per square inch, correlations between colors six inches apart, and the number of syllables in the artist's name. Facts *about* a work cannot be copyrighted and are not part of the expressive content of a work.[59]

This distinction between factual and expressive content was made clear by *Feist Publications, Inc. v. Rural Telephone Service Co*.[60] In *Feist,* Rural Telephone Service Co. (Rural) created a telephone directory that included listings for its customers, while Feist Publications, Inc. (Feist) was a company that specialized in producing area-wide telephone directories. Feist used Rural's listings without permission in its own directory, resulting in Rural suing Feist for copyright infringement.

---

[58] In Doe v. GitHub Inc., the plaintiffs accuse the defendants of "distributing" the input code "to Copilot users as if it were created by Copilot." No. 3:22-cv-06823, at *6 (N.D. Cal. filed Nov. 3, 2022). Although the plaintiff's admitted that "Codex and Copilot do not retain copies of the materials they are trained on," they argue that "[i]n practice, however, the Output is often a near-identical reproduction of code from the training data." *Id.* at *15. Likewise, in Anderson v. Stability AI LTD., the plaintiffs argue that "[t]hese 'new' images are based entirely on the Training Images and are derivative works of the particular images Stable Diffusion draws from when assembling a given output." No. 3:23-cv-00201, at *3 (N.D. Cal. filed Jan. 13, 2023). This leads them to conclude that the AI tool "is merely a complex collage tool." *Id.*

[59] "[C]opyright's idea/expression dichotomy strikes a definitional balance between the First Amendment and the Copyright Act by permitting free communication of facts while still protecting an author's expression. No author may copyright his ideas or the facts he narrates." Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 556 (1985) (internal citations omitted).

[60] 499 U.S. 340 (1991).

The key issue in the case was whether Rural's telephone directory was eligible for copyright protection. The Supreme Court held that the directory was not protected by copyright because it lacked the necessary originality and creativity. The *Feist* court emphasized that facts, such as names, addresses, and phone numbers, are not copyrightable because they are discovered rather than created, and thus do not meet the originality requirement. The court noted that *compilations* of facts may be copyrightable, but only those compilations that show sufficient human creativity.

Though similar to the phone books in *Feist,* ML training is even further from infringement in that the factual content recorded in the model is generated by the training process. In *Feist*, the factual content was directly copied from Rural's phone book. But in ML training, the statistical probabilities associated with each input are not part of the work at all. They are generated in response to the "predict" and "adjust" phases of the training process. This would be like if Feist recorded measurements such as the number of businesses associated with each letter, the correlation of phone numbers with names, and other similar facts, and then published those facts without publishing any part of the phone book itself. If the court was unwilling to find that straightforward facts violated copyright, then the argument for abstract and newly generated facts like the facts recorded in ML models is even stronger. Ultimately, whether abstract or straightforward, the rule remains that facts are not copyrightable.

Even viewing the model as a whole, the statistical measurements within a model are not selected due to any human creativity or judgment. It is the computer process that identifies correlations and records the facts. Moreover, humans are currently unable to even

understand the connection between any particular weight in the model and any fact observed during training.

One case that moves towards addressing "abstract facts", is *New York Mercantile Exchange, Inc. v. IntercontinentalExchange, Inc* which focused on the issue of copyright protection for market settlement prices.[61]

New York Mercantile Exchange, Inc. (NYMEX) sued IntercontinentalExchange, Inc. (ICE), claiming that ICE had infringed on NYMEX's copyrights by republishing its market settlement prices without authorization. NYMEX argued that the settlement prices were original and creative works deserving of copyright protection.

The Second Circuit disagreed with NYMEX's argument. The court held that the settlement prices were not copyrightable because they were factual information and not original expressions. The court reasoned that the prices were determined by an objective process involving the exchange of bids and offers and thus did not possess the requisite level of creativity and originality required for copyright protection. Like the measurements taken as part of the ML training process, the bid and offer prices in *New York Mercantile* were independent facts, albeit difficult to extract, available for anyone to use.

As the *Feist* court noted:

[F]acts do not owe their origin to an act of authorship. The distinction is one between creation and discovery: The first person to find and report a particular fact has not created the fact; he or she has merely discovered its existence. To borrow from Burrow-Giles, one who discovers a fact is not its "maker" or "originator." The discoverer merely finds and

[61] 497 F.3d 109 (2d Cir. 2007).

records. Census takers, for example, do not "create" the population figures that emerge from their efforts; in a sense, they copy these figures from the world around them.[62]

Just like the census taker in the example given by the Supreme Court, the ML model records only facts in the form of statistical probabilities. These facts are available for anyone, or any process, to copy and use. And just as there is no copyrightable expression in a mechanistic set of measurements about a work, there is no expression copied *from* the work to make such a set of facts.

The Use in ML Training is a Research Purpose

The third consideration relevant to the character and use of the work is the special deference given to research purposes by the Copyright Act.[63] This consideration acknowledges the importance of promoting the progress of knowledge and fostering innovation, which are both key objectives of copyright law. When copyrighted material is used for research purposes, courts are more likely to find that it is fair use, as it supports the advancement of knowledge and serves the greater public good.

The facts developed and recorded in the model through the ML training process are a type of research output. Not only are the models a research topic in the computer science context, but the statistical insights encoded in ML models themselves have produced new

[62] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 347 (1991) (internal citations omitted).
[63] 117 U.S.C. §107 ("reproduction in copies…for purposes such as…teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright.").

insights into domains like linguistics[64] and art history[65] Scientists writing in the journal *Nature* said:

AI can act as an instrument revealing properties of a physical system that are otherwise difficult or even impossible to probe. Humans then lift these insights to scientific understanding. Second, AI can act as a source of inspiration for new concepts and ideas that are subsequently understood and generalized by human scientists. Third, AI acts as an agent of understanding. AI reaches new scientific insight and — importantly — can transfer it to human researchers.[66]

First Factor Arguments Against Fair Use

Some considerations under the first fair use factor weigh against a finding of fair use. The first is the commercial nature of much ML training activity. As described in the copyright statute, the purpose and character of the use includes "whether such use is of a commercial nature or is for nonprofit educational purposes."[67] There is substantial legitimate academic and nonprofit activity that includes making ML models and generating

---

[64] Steven Piantadosi, *Modern language models refute Chomsky's approach to langauge*, LINGBUZZ (Mar. 2023) https://lingbuzz.net/lingbuzz/007180 ("Modern machine learning has subverted and bypassed the entire theoretical framework of Chomsky's approach, including its core claims to particular insights, principles, structures, and processes. I describe the sense in which modern language models implement genuine theories of language, including representations of syntactic and semantic structure.").

[65] Will Fenstermaker, *How Artificial Intelligence Sees Art History*, MET (Feb. 4, 2019), https://www.metmuseum.org/perspectives/articles/2019/2/artificial-intelligence-machine-learning-art-authorship ("Machine learning, he said, makes it possible to begin visualizing the diversity and complexity of artistic creation. . . . But by harnessing artificial intelligence, it's possible to envision artworks that may have existed, based on our knowledge of artworks that we know to exist, and in that way gain a fuller, more complete understanding of visual culture."); Ahmed Elgammel, *The Shape of Art History in the Eyes of the Machine*, 32 Thirty-Second AAAI Conf. A.I. 2183, 2183 (2018) ("[T]he machine can learn an internal representation encoding discriminative features through its visual analysis . . . he learned representations also consistently highlighted certain artists as the extreme distinctive representative of their styles, which quantitatively confirms art historian observations.").

[66] Mario Krenn et al., *On scientific understanding with artificial intelligence*, 4 NATURE REV. PHYSICS 761, 761 (2022).

[67] 17 U.S.C. §107(1).

content using them,[68] but in this hypothetical there is a clear commercial incentive for the ML activity.

As expressed by the Supreme Court in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith et al*., ("*Warhol*")[69] "the first fair use factor instead focuses on whether an allegedly infringing use has a further purpose or different character, which is a matter of degree, and the degree of difference must be weighed against other considerations, like commercialism."[70]

In *Warhol*, Goldsmith took a photograph of Prince in 1981 and Warhol created a series of silkscreen prints of the photograph in 1984.[71] The prints were part of Warhol's "Prince Series," which consisted of 40 portraits of Prince. Goldsmith argued that Warhol's prints infringed her copyright in the original photograph. She argued that the prints were "substantially similar" to the original photograph and as such were derivative of the original work.[72]

The Andy Warhol Foundation argued that the prints were protected by fair use. The foundation argued that the prints were transformative because they had a different

[68] *See, e.g.* the machine learning category of arxiv,org, a repository of scholarly papers dealing with machine learning, at https://arxiv.org/list/cs.LG/recent.
[69] Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith et al.., ____ U.S. ____, ____ S. Ct. _____(2023), *slip. op available at* https://www.supremecourt.gov/opinions/22pdf/21-869_87ad.pdf.
[70] *Id.* at 12, citing Campbell v. Acuff-Rose Music, Inc., 510 U. S. 569, 579 (1994). *But see* Google LLC v. Oracle Am., Inc., 141 S. Ct. 1183 (2021) ("There is no doubt that a finding that copying was not commercial in nature tips the scales in favor of fair use. But the inverse is not necessarily true, as many common fair uses are indisputably commercial. For instance, the text of §107 includes examples like "news reporting," which is often done for commercial profit. So even though Google's use was a commercial endeavor—a fact no party disputed, see 886 F. 3d, at 1197—that is not dispositive of the first factor….")
[71] *Warhol*, slip. op. at 3-4.
[72] *Id.*, at 8-9, 11.

"meaning and message" than Goldsmith's original photo and were used for a different purpose, and as such did not harm Goldsmith's market for the original photograph.[73]

The Supreme Court granted *certiorari* on the single issue of whether the first fair use factor weighed in favor of fair use in the specific circumstance of the Andy Warhol Foundation's licensing of "Orange Prince" to Condé Nast for use in a magazine.[74] The *Warhol* court then found that AWF's licensing of Orange Prince was not justified by the first fair use factor.[75]

In general, commercial use of a derivative work makes a finding of fair use less likely. However, the distinctions between the works at issue in *Warhol* and how ML models are trained tends to reduce the significance of the commercial nature of some ML model building. As expressed by the *Warhol* court, "[t]he fair use provision, and the first factor in particular, requires an analysis of the specific "use" of a copyrighted work."[76] The court found that only "AWF's commercial licensing of Orange Prince" was unjustified, and in particular that the Court "expresses no opinion as to the *creation*… of the Prince Series works."[77]

In contrast, the differences between ML models and the works they are trained on are so stark that there is no reasonable comparison between them. The *Warhol* court found that both Orange Prince and Goldsmith's original photograph were licensed for magazine covers, showing that the two works were substitutes in that market. However, an ML model

[73] *Id.* at 9-10.
[74] *Id*. at 11-12.
[75] *Id.* at 2.
[76] *Id*. at 20.
[77] *Id.* at 21 (emphasis added).

is not viewable or intelligible in the same way as the works used to train the model. For example, Goldsmith's photo could be an input to an ML training procedure, but the ML model trained in part on Goldsmith's photo could not be licensed to replace the original photo in any circumstance. It is conceivable that an image later generated using the model could possibly be infringing, but the model itself is distinct.

The First Fair Use Factor Weighs Heavily in Favor of Fair Use

For any (and all) of the reasons listed above, the use of copyrighted works for ML training leans heavily–almost decisively–in favor of fair use.

In the recent case *Google LLC v. Oracle America, Inc.*,[78] the Supreme Court highlighted the importance of transformativeness. In *Google v. Oracle*, Google exactly copied portions of Oracle's copyrighted source code. The copied source code had the same meaning and message. But the Supreme Court found that Google's use was transformative because of the way in which it drove the creation of new, independent works. "To the extent that Google used parts of the Sun Java API to create a new platform that could be readily used by programmers, its use was consistent with that creative 'progress' that is the basic constitutional objective of copyright itself."[79]

Generative ML models are *much more* transformative than the copied source code at issue in *Google v. Oracle*. It is much more transformative than search indexes at issue in the *Authors Guild* cases, or any of the other cases that found fair use. People can and do use generative ML models to facilitate the almost unlimited generation of *new* works.

[78] *Id.*
[79] *Id.*

Generative ML models make artistic creation accessible to a broad portion of the population–and it is evident that new works are being created every hour of every day. This fulfills the "basic constitutional purpose" of copyright to an unprecedented degree.

The Nature of the Copyrighted Work

The second fair use factor is the nature of the copyrighted work. This factor does not influence the fair use analysis either way. None of the types of works that could be used for training receive any special favor or analysis. To the ML application, the exact type of content used for training is irrelevant; the model only sees a series of numbers. All types of works are treated equivalently. Thus this factor does not bear any weight in the analysis. This is especially true when, as here, the creative work is used for a transformative purpose.[80]

The Amount or Substantiality of the Portion Used

The third factor in the fair use analysis is whether the amount of copying exceeded what was necessary and if it was excessive. There are no strict rules on how much of a copyrighted work can be copied while still being considered fair use.[81] The permissible extent of copying depends on the purpose and character of the use. Excessive copying is copying anything "more" than what is reasonably "necessary."[82] In some cases, copying

[80] Cariou v. Prince, 714 F.3d 694, 710 (2d Cir. 2013) (quoting Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 612 (2d Cir. 2006)).

[81] Maxtone-Graham v. Burtchaell, 803 F.2d 1253, 1263 (2d Cir. 1986). "[T]he extent of permissible copying varies with the purpose and character of the use." Campbell, 510 U.S. at 586-87. "The crux of the inquiry is whether "no more was taken than necessary." Id. at 589.

[82] *See* Harper & Row v. Nation Enterprises, 471 U.S. 539, (1985).

the entire work might be necessary, and in such instances, this factor doesn't weigh against a finding of fair use.

In the case of ML models, the purpose is to get as wide an exposure to different types of inputs as possible. Further, it isn't reasonable to view (or train on) just *part* of a picture or *part* of an article. The situation is similar to *Authors Guild, Inc. v. Hathitrust.* Just as it was reasonable in *Hathitrust* to read entire books to create a full-text index, using entire works is a reasonable way to train an ML model.[83] Since using the entire works is reasonably necessary to enable ML model training, the copying is not excessive. Accordingly, the third fair use factor does not incline the fair use analysis either way.

The Effect on the Market

The last factor in the fair use analysis is how the use affects the market for the original work. As with the first fair use factor, this factor weighs heavily in favor of fair use.

The fourth fair use factor properly addresses possible markets—but it does not include *all* the hypothetical markets that copyright holders could pursue. Those markets may exist, but the possible "market harm" is the extent to which the result of the copying serves as a substitute for the original work.[84] As stated by the Supreme Court in *Campbell v. Acuff-Rose Music*: "[T]he only harm to derivatives that need concern us, as discussed

[83] Authors Guild v. Hathitrust at 210.

[84] "Even when an entire copyrighted work was recorded, the District Court regarded the copying as fair use because there is no accompanying reduction in the market for plaintiff's original work." Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 425-26, 104 S. Ct. 774, 780 (1984); Fair use depends on "the likelihood that the parody may serve as a market substitute for the original", Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 586, 114 S. Ct. 1164, 1175 (1994).

above, is the harm of market substitution."[85] Taking the example of the *Warhol* court, both the Goldsmith photograph and the Warhol print were licensed for the same purpose—use in a magazine. Other possible uses, like display of the Prince Series in a museum, were not examined.

Trained ML models and ML applications are wholly different types of goods than the inputs that they trained on. There is no possible market substitution between the ML model and any particular input it is trained on. The ML model is useless as an artwork, song, poem, or as any other type of creative work. As described above, anyone looking at an ML model would only see a "gigantic matrix of numbers" inscrutable to any process but the ML application itself.[86]

Even if a separate recognizable market for training ML models develops, it is hard to argue that training an ML model would have a significant effect on the market for any one particular work used in the model training. What matters in ML model training is volume. Models are trained on millions of works. The contribution of each individual work to the model weights is so small as to be nearly imperceptible, if it can be measured at all.

Using Copyrighted Works to Train an ML Model is Fair Use

From the analysis above, it becomes clear that ML training is "quintessential" fair use. When ML model training is examined with the correct factual background, the strength of its legality surpasses that of even the most obvious and well-known cases.

[85] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 593, 114 S. Ct. 1164, 1178 (1994).

[86] *Supra,* "Defining the Model." Also see . LeCun, "My take on Ali Rahimi's "Test of Time" award talk at NIPS," 2017, available at https://www2.isye.gatech.edu/~tzhao80/Yann_Response.pdf ("The engineering artifacts have almost always preceded the theoretical understanding").

There is no need to posit a special exception for "fair learning"[87] to address ML model training. Existing case law convincingly makes the case that ML model training is fair use.

It is undisputed that copyrighted works are necessary for many types of ML training. But as stated by the *Feist* court, "[t]he primary objective of copyright is not to reward the labor of authors, but to promote the Progress of Science and useful Arts. To this end, copyright assures authors the right to their original expression, but encourages others to build freely upon the ideas and information conveyed by a work."[88] This is what ML models enable people to do.

This analysis has been presented in the context of a hypothetical lawsuit against a generative image service using the Stable Diffusion model. The specifics of the German training location are tied to this hypothetical. All the analysis regarding fair use, however, has been agnostic to the type of input used to train the ML model. This is because ML models can't see or appreciate the expression that is central to copyright. ML models are the classic "literate robot"[89]–an automated process that courts have found to be fair use because the copyright-protected expression is never exposed to a human viewer.

3. Providing a Generative Service Using Machine Learning Models is Fair Use

The lawsuits against ML applications have targeted the training process as part of the complaint due to the copying that occurs as part of the training process.[90] But it is

[87] *See, e.g.,* Lemley, Mark A. and Casey, Bryan, *Fair Learning*, 99 TEX. L. REV. 4 (2020).
[88] Feist Publ'ns Inc., 499 U.S. at 349-50 (internal quotation marks and citations omitted).
[89] James Grimmelmann, *Copyright for Literate Robots*, 101 IOWA L. REV. 657 (2016).
[90] "Stability scraped, and thereby copied over five billion images from websites as the Training Images used as training data for Stable Diffusion." Anderson v. Stability., No. 3:23-cv-00201 (N.D. Cal. filed Jan.

typically not the *inputs* to ML models that are the real source of disputes, but the generative *outputs* that are the most concerning to artists and authors.[91] The analysis of the generative aspects primarily revolves around two questions: first, if the output of the ML application is a reproduction or derivative work of one or more of the inputs, and second, if the output of the ML application implicates a copyright interest in one or more works used as inputs, then is providing the ML application itself inducing infringement.

Generating Content Using an ML Application

The threshold question for any copyright infringement is whether a particular copyrighted work has been copied. Looking specifically at the Stable Diffusion-based application from the hypothetical, there are two scenarios in which it is possible to generate works that would clearly infringe a particular copyrighted work.

Direct Reproduction of Inputs

The first scenario was discussed in the context of overtraining and memorization above.[92] A group of AI researchers from Google, DeepMind, UC Berkeley, Princeton, and

13, 2023) at *15; "Stability AI has copied more than 12 million photographs from Getty Images' collection, along with the associated captions and metadata, without permission from or compensation to Getty Images, as part of its efforts to build a competing business." Getty Images, Inc. v. Stability AI, Inc., No. 1:23-cv-00135-UNA (D. Del. filed Feb. 3, 2023) at *1.

[91] "These resulting derived images compete in the marketplace with the original images. Until now, when a purchaser seeks a new image "in the style" of a given artist, they must pay to commission or license an original image from that artist." Anderson v. Stability at *1; "Stability AI now competes directly with Getty Images by marketing Stable Diffusion and its DreamStudio interface to those seeking creative imagery…." Getty Images, v. Stability AI at *3; *see generally Artists decry use of AI-generated art,* THE INDEPENDENT (Dec. 10, 2022), available at "https://www.independent.co.uk/news/world/americas/ai-art-lensa-magic-avatar-b2242891.html.

[92] Supra at "Overtraining and Memorization."

ETH Zurich developed a process by which they could extract almost visually identical copies of one hundred nine of the inputs used in the Stable Diffusion dataset.[93]

It is instructive, however, to understand the process used by the researchers to extract these matches. For the images extracted from Stable Diffusion, the researchers had preexisting knowledge of exactly how the ML model was trained. They selected the 350,000 most duplicated images in the dataset–i.e., the images most likely to suffer from overtraining–and the exact terms associated with those images in the model.[94] With this inside knowledge, they prompted the generation of five hundred images for each of the 350,000 target images using exactly matching terms–175 million total generated images.[95] Each of these 175 million images was inspected by an automated process for similarity to the target image.[96] The result was the production of one hundred nine successfully visually similar images-- an extremely minor occurrence, about three percent.[97] Not only that, but to even retrieve, or "discover" that this replication occurred, required immense effort and searching. With some understatement, the researchers commented that successfully extracting duplicated images from Stable Diffusion was "computationally expensive."[98]

---

[93] Nicholas Carlini, et al., *Extracting Training Data from Diffusion Models*, arxiv.org, (2023) (The researchers were also able to force a different application using a much smaller dataset (60,000 images) to regenerate about 1250 of its inputs by retrying the process a million times and comparing every generated image to every input).

[94] *Id.* at 5.

[95] *Id.*

[96] *Id.* at 4.

[97] Keep in mind that this was done using a dataset of the most duplicated images. If the process were repeated with the other images used by the model, we would expect the number to be much smaller, perhaps even non-existent. Again, scenarios where duplication becomes a possibility, known as "overtraining," are avoided by ML designers.

[98] *Id.* at 5.

One reason why it was so difficult to find duplicated images is because the "duplicates" were not identical to the inputs. Though very similar, the duplicates were "degraded", meaning they had noise or trivial differences distinguishing them from the inputs.

Thus, while direct reproduction of inputs due to memorization is possible, it is generally rare. Researchers have been able to provoke models to generate outputs similar to inputs,[99] but direct copying is also unlikely to be the goal of users due to the degraded form of the outputs. If anyone wanted a pristine copy of an image (or any other type of input), they would simply make a copy, rather than use an ML application to generate a poor reproduction.[100]

Finally, direct reproduction becomes more unlikely and more difficult as ML models get trained on more inputs, and the training sets are filtered to remove duplicate inputs. The larger and more diverse the training set, the greater the capability and likelihood that a model will generate wholly new works.

[99] In the case of generative text applications, recent experiments have found that they may be more likely to reproduce portions of their inputs. As with generative image models this is likely the result of inadvertent overtraining due to the duplication of inputs in the training set. See, e.g., Henderson et al., *Foundation Models and Fair Use*, available at https://deliverypdf.ssrn.com/delivery.php?ID=394013103113027029124025108118071109050063050068079069007115006009071030107023119009034102098026110059062000090125091003002065045007060077040003019098086127108048012022119125003087001010029004085015086093100100085079001026010066006098025065012020&EXT=pdf. Updated training methods that avoid duplicated would likely make the reproduction of inputs less likely.

[100] Intent doesn't matter for copyright infringement. However, the likelihood of a course of action is relevant to the fair use analysis.

Character Copyright

The second scenario has to do with copyrighted characters. Recognition of characters as independently copyrightable works emerged in 1930 with the case of *Nichols v. Universal Pictures*.[101] In *Nichols*, the Second Circuit Court of Appeals denied protection to the plaintiff's characters because they were not "distinctly delineated" but rather poorly developed. The characters, a Jewish gentleman and the poor Irish Catholic girl he loved, were considered mere 'prototypes.' Judge Hand stated that the less developed a character is, the less copyrightable it becomes.[102] In denying the copyrightability of the poorly-developed characters at issue in *Nichols,* the court left open the possibility that well-delineated characters could be copyrighted separate from any of the works in which they appear. Later courts applying this test found that the character of Tarzan was found to be "sufficiently delineated" and protected by copyright.[103] Similarly, Superman's character was deemed well-delineated due to its original literary expressions and incidents, thus deserving copyright protection.[104]

In the context of images, reproductions of well-known characters in new, independently created situations have been found to infringe a character's copyright. In *Walt Disney Productions v. Air Pirates*,[105] the court examined a series of comic books that depicted famous Disney characters engaging in counter-cultural activities, including

[101] 45 F.2d 119 (2d Cir. 1930), cert. denied, 282 U.S. 902 (1931).
[102] Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930).
[103] Burroughs v. Metro-Goldwyn-Mayer, Inc., 683 F.2d 610 (2d Cir. 1982).
[104] Detective Comics, Inc. v. Burns Publications, 111 F.2d 432; 434 (2d Cir. 1940).
[105] 581 F.2d 751 (9th Cir. 1978).

promiscuity and drug use.[106] These unauthorized comic books parodied and subverted the wholesome image of the Disney characters, which led to the legal dispute over copyright infringement. The court held that the Disney characters were copyrightable, and that the copyright was infringed, based on the grounds that the comic-book characters had distinctive "physical as well as conceptual qualities" that was "likely to contain some unique elements of expression."

Unlike every other type of copyrighted work recognized by the courts, copyrighted characters are not limited to a single expression. As a consequence, people using ML applications can create new works that show well-known, possibly copyrighted characters, in new situations. This is a particular risk for image-generating ML applications, because it seems likely that just as an ML model learns to emulate a van Gogh painting, it might learn to generate a facsimile of a cartoon character like Superman or Iron Man. Even though the model would likely never reproduce an existing image of the character, producing new scenes with "old" characters might be infringing because of the visually distinctive markings that are associated with these types of characters. Because of this possibility, the question now becomes whether providing these technologies would "induce infringement."[107]

[106] *Id*. at 753.

[107] For simplicity, this article assumes that an image incorporating a copyrighted character would be infringing. Nevertheless, an image that included a character might still be fair use for other reasons specific to that work.

The Significant Noninfringing Uses of ML Applications

The analysis of whether a product might induce the infringement of another's copyright is driven by whether the product has "significant noninfringing uses" and whether the product is marketed to users as a means to infringe copyright.

The "significant noninfringing uses" doctrine originated from *Sony Corp. v. Universal City Studios* discussed above.[108] In *Sony*, the accused product was Sony's Betamax video cassette recorder (VCR). It was undisputed by both sides in *Sony* that Sony's VCRs were capable of making infringing copies of movies or other material. However, the court found that if a product had substantial noninfringing uses, the manufacturer would not be liable for contributory copyright infringement. In the case of the VCR, making copies for private, non-commercial use was sufficient.[109] The Court reasoned that stifling the distribution of products with legitimate uses would impede technological progress and undermine the goals of copyright law–to promote the progress of science and useful arts.

In *MGM Studios v. Grokster*,[110] the Supreme Court clarified the doctrine by introducing the concept of "inducement" to the analysis of secondary liability. The *Grokster* Court held that even if a technology has substantial noninfringing uses, its

[108] 464 U.S. 417 (1984).

[109] "The question is thus whether the Betamax is capable of commercially significant non-infringing uses. In order to resolve that question, we need not explore all the different potential uses of the machine and determine whether or not they would constitute infringement. Rather, we need only consider whether on the basis of the facts as found by the District Court a significant number of them would be non-infringing. "Moreover, in order to resolve this case we need not give precise content to the question of how much use is commercially significant.", Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 442, 104 S. Ct. 774, 789 (1984).

[110] 545 U.S. 913 (2005).

distributor may still be liable for copyright infringement if the service actively induces users to infringe copyrights.

In the context of generative ML applications, the primary purpose (and the primary use) is the generation of new works. Compare with copy machines, which have significant noninfringing uses even though they are designed to make it easy to copy things.

In contrast, generative ML is defined by its ability to *generate* new things; it is a poor copyist. While it is possible to generate infringing works using such applications, the overwhelming majority of users generate original art, original text, or original code. This is not just a "significant noninfringing purpose," it is in furtherance of the purposes of copyright. Assuming marketing consistent with the generative aspect of generative ML, there should be no secondary liability on the part of a party hosting a generative ML service, even if it can possibly be used to create possibly-infringing works.

Generating Works "In the Style of" Particular Artists

One of the most controversial aspects of generative ML models is the ability to create works "in the style of" a known artist.[111] Prompts that include specific artists' names can generate works that are strongly reminiscent of that artist's style. These ML-generated works can and likely do compete with the original artists in the marketplace.[112]

[111] *See, e.g., 'In the style of': why AI art needs to address named artists as prompts*, BYTESIDE, available at https://www.byteside.com/2022/09/ai-art-named-artists-monet-picasso-rutkowski/; *Is A.I. Art Stealing from Artists?,* THE NEW YORKER, available at https://www.newyorker.com/culture/infinite-scroll/is-ai-art-stealing-from-artists.

[112] *See, e.g.* "Stability AI Competes Commercially with Getty Images," Getty Images, Inc. v. Stability AI, Inc., at *15 (D. Del. filed Feb. 3, 2023).

The difficulty for artists is that "style," standing alone, has not generally been found to be copyrightable.[113] That said, style is not completely divorced from expression. Courts have found that a copied "style" is more likely to indicate that one work was copied from another. For example in *Steinberg v. Columbia Pictures*,[114] the court said: "Even at first glance, one can see the striking stylistic relationship between the posters, and since style is one ingredient of 'expression,' this relationship is significant." However, this statement was in the context of both works being substantially similar, meaning that courts only consider style if the works are first substantially similar enough to be considered substitutes. For example, in *Steinberg* the court first noted that "[b]oth illustrations represent a bird's eye view across the edge of Manhattan…. Both depict approximately four city blocks in detail and become increasingly minimalist as the design recedes into the background."[115] After observing this substantial similarity, only then did the court consider style: "Both use the device of a narrow band of blue wash across the top of the poster to represent the sky, and both delineate the horizon with a band of primary red."[116] In other

[113] "[o]f course, the idea of animal styled duffle bags would not be protectible under copyright law.", Wildlife Express Corp. v. Carol Wright Sales, Inc., 418 F.3d 502, 510 (7th Cir. 1994); "Our decision does not grant license to copyright a musical style or "groove."", (Williams v. Gaye, 895 F.3d 1106, 1138 (9th Cir. 2018)); "[Plaintiff] wants to copyright a style…, and no sensible reading of the 1976 Act permits that step…. [The accused work] Liza does, however, convey an impression similar to Mara's….[By] contending that Liza infringes the Mara copyright, [Plaintiff] demonstrates that its claim embraces an aesthetic style rather than a precise set of features.", Pivot Point Int'l v. Charlene Prods., Inc., F. Supp. 2d 828 (N.D. Ill. 2001) (decision upheld on remand); *but see* Jacobs v. Robitaille, 406 F. Supp. 1145, (D.N.H. 1976) (holding that copying "style" may result in unfair competition).

[114] Steinberg v. Columbia Pictures Indus., 663 F. Supp. 706 (S.D.N.Y. 1987).

[115] *Id.* at 712.

[116] *Id. See also* Ford Motor Co. v. B & H Supply, Inc., 646 F. Supp. 975 (D. Minn. 1986) (describing the similar "style" as an element in finding two works substantially the same).

words, there must be substantial similarity, not just a similarity in "style," to find that one work infringes another.

When separated from specific copied elements between two works, the concept of "style" has been found to lie more in the realm of an *idea* than an expression. For example, the court in *Dave Grossman Designs v. Bortin*[117] stated:

> The law of copyright is clear that only specific expressions of an idea may be copyrighted, that other parties may copy that idea, but that other parties may not copy that specific expression of the idea or portions thereof. For example, Picasso may be entitled to a copyright on his portrait of three women painted in his Cubist motif. Any artist, however, may paint a picture of any subject in the Cubist motif, including a portrait of three women, and not violate Picasso's copyright so long as the second artist does not substantially copy Picasso's specific expression of his idea.[118]

This distinction matters because from a copyright perspective, the relevant market–fourth fair use factor–is the market for a *particular work*, not for an artist's work *in general*.[119] The case law is clear: the copying of an artist's distinctive style in the context of a new image is not an infringement of the artist's copyright in any particular work.[120] Looking at other types of works, doing things "in the style of" another artist is even more attenuated. There is no copyrightable interest in the written style of a particular author, nor of the general style of a musical artist. There must always be specific copied expression. Thus, ML applications that generate new works "in the style of" a particular artist are not

[117] Dave Grossman Designs, Inc. v. Bortin, 347 F. Supp. 1150 (N.D. Ill. 1972).
[118] *Id.* at 1156.
[119] "[W]hen a commercial use amounts to mere duplication of the entirety of an original, it clearly "supersede[s] the objects," Folsom v. Marsh, supra, at 348, of the original and serves as a market replacement for it, making it likely that cognizable market harm to the original will occur." Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 591, 114 S. Ct. 1164, 1177 (1994).
[120] An artist or author may have causes of action other than copyright, including trademark liability or potential right of publicity claims. However, those are not the focus of this article.

infringing–and to the extent that the result is the generation of new works, images "in the style of" another artist further copyright's overall purpose.

4. Conclusion

There are many discussions of how ML–especially generative ML–will change society and the law. The effects of machine learning and ML applications are sure to create upheaval. As the applications of machine learning continue to expand and evolve, it is crucial for legal frameworks to adapt and ensure that innovation is not stifled while still maintaining the core objectives of copyright law—promoting the progress of arts and sciences while protecting the rights of creators.

By delving into the fundamentals of machine learning, including the training process and the generation of new works, this analysis drew parallels and distinctions between ML and previously scrutinized technologies in the context of copyright law…. While this article has focused on image-generating ML tools, the analysis applies to all the various types of generative AI. The case law overwhelmingly supports the conclusion that constructing and utilizing generative ML models is allowable under US copyright law.

FOR-PROFIT INCARCERATION: AN EVALUATION OF THE RELIGIOUS LAND USE AND INSTITUTIONALIZED PERSONS ACT IN THE ERA OF PRIVATE PRISON BUSINESS MODELS

Anthony J. Papageorgiou[1]

I. Introduction

The First Amendment of the United States Constitution requires Congress to make no law respecting an establishment of religion or prohibit the free exercise thereof.[2] Congress has taken steps to reinforce this protection through its passage of the Religious Land Use and Institutionalized Persons Act (RLUIPA).[3] Under federal law, a prison or jail cannot substantially burden a prisoner's exercise of their religion unless it can demonstrate that it has a compelling interest that cannot be achieved through any other, less restrictive means.[4]

The law protects these rights for a good reason.[5] Available empirical evidence suggests an inverse relationship between religion and crime.[6] According to over 40 years of empirical research summarizing the relationship between religion and crime, findings indicate that religion decreases propensities for criminal behavior.[7] While in prison, places of worship often act as a safe haven for inmates, and where the safety threats of prison life are lessened.[8] Involvement in religion also exposes a prisoner to fewer problems of prison life, merely by socialization with other religious inmates.[9]

[1] Special thank you to John, Irene, and Georgia Papageorgiou for their limitless support and encouragement.
[2] U.S. CONST. amend. I.
[3] *Religious Freedom in Prison*, AM. CIV. LIBERTIES UNION, https://www.aclu.org/issues/prisoners-rights/civil-liberties-prison/religious-freedom-prison add (last visited October 24, 2022)
[4] *Id.*
[5] *Id.*
[6] Travis Hirschi & Rodney Stark, *Hellfire and Delinquency,* 17 SOC. PROBS. 202 (1969).
[7] *Id.*
[8] Todd R. Clear, et al., *Does Involvement in Religion Help Prisoners Adjust to Prison?*, THE NAT'L COUNCIL OF CRIME AND DELINQ., 1992, https://www.evidentchange.org/sites/default/files/publication_pdf/religion-and-prisoners.pdf.
[9] *Id.* at 6.

Following the formalization and centralization of prisons, courts and legislatures have tangled with the idea of free religious exercise among prisoners, often overreaching and giving broad discretion to correctional authorities.[10] This is done in an effort to improve security and promote the efficient day-to-day operations of confinement.[11] Unfortunately, allowing such ambiguous and broad discretionary power to correctional staff has negative consequences about future inmate recidivism and the overall quality of life within prison walls.[12] These harms are compounded by the private prison business model primarily concerned with bottom-line profit. This is because the United States correctional system, at large, is generally overworked, understaffed, and scrambles to provide its inmates with even their most basic human needs.[13]

Private prisons business models add an additional obstacle when financial incentives often conflict with providing inmates more resources. If the legislature does not take any action on its own, courts should reevaluate their approach and place an objective, reasonable person standard regarding a prisoner's expectations and fundamental rights to practice religion freely. It is vital that courts embrace this approach and encourage a more encompassing variety of religious practices to flow within prisons. This is an evidence-based approach to providing a safer and more secure prison environment while ultimately combating recidivism.[14]

II. Discussion

A. Strict Scrutiny Standard

[10] 42 U.S.C. § 2000cc-1(a)(1)-(2) (stating that "no government shall impose a substantial burden on the religious exercise of a person residing in or confined to an institution unless the burden furthers a compelling interest and does so by the least restrictive means").
[11] *Holt v. Hobbs*, 574 U.S. 352, 362 (2015) (discussing that correctional authorities have a broadly formulated interest in upholding prison safety and security).
[12] Clear, et al., *supra* note 8, at 6.
[13] *Furgess v. Pa. Dep't of Corr.,* 933 F.3d 285, 287 (3d Cir. 2019) (where an inmate was unable to shower for three months because prison staff did not provide him with a handicapped-accessible shower facility).

[14] *Id.*

In United States Constitutional law, when a court finds that a law infringes someone's fundamental constitutional right, it ordinarily applies what is called a strict scrutiny standard in an attempt to uphold the law or policy constitutionally valid.[15] The infringing law is permissible only if the government can demonstrate in court that the law or regulation is necessary to achieve a "compelling state interest."[16] The government must also demonstrate that the law is "narrowly tailored" to achieve the compelling purpose and must have used the "least restrictive means" to achieve the purpose.[17] Failure to show any of these conditions will likely result in a judge striking down a law as unconstitutional.[18]

The standard is the highest and most stringent standard of judicial review and is part of the levels of judicial scrutiny that courts use to determine whether a constitutional right or principle should give way to the government's interest against observance of the principle.[19]

B. The Evolution of the Free Exercise Clause in Prisons

By the mid-1900s, a series of Supreme Court rulings about general restrictions on religious freedoms concerned Congress.[20] In *Sherbert v. Verner* (1963), states were required to provide unemployment compensation to individuals who lost their jobs because the jobs conflicted with their Sabbaths.[21] In *Wisconsin v. Yoder* (1972), the Court upheld the right of Amish parents to withdraw their children from schools after the eighth grade.[22] Both cases required that states pass

[15] Richard H. Fallon, Jr., *STRICT JUDICIAL SCRUTINY*, 54 UCLA L. Rev. 1267, 1268 (2007).
[16] *Id*. at 1267
[17] *Id*. at 1326.
[18] *Id.* at 1315-16.
[19] Roy G. Spece, Jr. & David Yokum, *Scrutinizing Strict Scrutiny*, 40 VT. L. REV. 285, 288 (2015).
[20] John R. Vile & David L. Hudson, *Religious Land Use and Institutionalized Persons Act of 2000*, THE FIRST AMENDMENT ENCYCLOPEDIA, (Sep. 2017), https://www.mtsu.edu/first-amendment/article/1093/religious-land-use-and-institutionalized-persons-act-of-2000.
[21] *Sherbert v. Verner*, 374 U.S. 398, 410 (1963).
[22] *Wisconsin v. Yoder*, 406 U.S. 205, 236 (1972).

a strict scrutiny, "compelling interest" test before imposing restrictions on religious entities or practices.[23]

Later, in examining a state law that did not provide unemployment compensation to Native Americans that had been fired from their jobs for ingesting peyote for religious reasons, the Court decided in *Employment Division, Department of Human Resources of Oregon v. Smith* (1990) that it did not have to apply such a stringent test in cases where laws of general applicability happened to fall with greater force on religion.[24]

In an attempt to overrule *Employment Division* and protect religious exercises, Congress, relying chiefly on its enforcement powers in Section 5 of the Fourteenth Amendment, required courts to apply the earlier *Sherbert* "compelling state interest" test when it adopted the Religious Freedom Restoration Act (RFRA) of 1993.[25] This lasted until the *City of Boerne v. Flores* decision of 1997.[26] There, the court struck down the RFRA as unconstitutional and its requirement that the state apply the strict scrutiny/compelling interest test to a case involving the zoning of a church.[27] The Court instead decided that Congress had exceeded its powers in attempting to dictate to state and local governments.[28] By exercising its remedial and preventive power to enforce a constitutional right under Section 5 of the Fourteenth Amendment, the court concluded that going forward, Congress is only to enact legislation that utilizes congruent and proportional means in achieving their legislative purpose.[29]

[23] Vile & Hudson, *supra* note 20.
[24] *Id*.
[25] *Id*.
[26] *Id.*; *City of Boerne v. Flores*, 521 U.S. 507, 536 (1997) (overturning the previous strict scrutiny, compelling interest test).
[27] *Flores*, 521 U.S. at 536.
[28] *Id*.
[29] *Id. at 533.*

In response, Congress relied on its spending power and its power over interstate commerce to adopt RLUIPA in 2000.[30] Then, in an establishment clause challenge in *Cutter v. Wilkinson* (2005), the Supreme Court upheld the portion of the act relative to prisoners.[31]

C. RLUIPA and Prisoner's Religious Rights in its Current Form

RLUIPA has had a significant impact on prisoners' religious liberty rights.[32] Prisoners have more statutory protection for their religious liberty rights under RLUIPA than under the standard Free Exercise Clause of the Constitution.[33] The primary example of this is evidenced in the Supreme Court's decision in *Holt v. Hobbs* (2015).[34] There, the Court unanimously ruled that Arkansas prison officials violated the religious-liberty rights of an inmate who was prohibited from growing a short beard.[35] Prison officials claimed security reasons for prohibiting the growing of such beards, but Justice Samuel A. Alito responded that such an argument was "hard to take seriously."[36]

A case that clearly defined the current jurisprudence on a prisoner's right to practice their religion freely is *Spratt v. Rhode Island Department of Corrections* (2007).[37] There, the court stated that a claim under RLUIPA includes four essential elements.[38] (1) That an institutionalized person's religious exercise has been burdened and (2) that the burden is substantial. On the first two elements, the plaintiff bears the burden of proof.[39] Once a plaintiff has established that his religious exercise has been substantially burdened, the onus shifts to the government to show (3)

[30] Vile & Hudson, *supra* note 15.
[31] *Cutter v. Wilkinson*, 544 U.S. 709, 726 (2005).
[32] Vile & Hudson, *supra* note 15.
[33] *Id.*
[34] *Holt v. Hobbs*, 574 U.S. 352, 370 (2015).
[35] *Id. at 369.*
[36] *Id. at 363-64.*
[37] *Spratt v. R.I. Dep't of Corr.,* 482 F.3d 33, 38 (1st Cir. 2007).
[38] *Id.*
[39] *Id.*

that the burden furthers a compelling governmental interest and (4) that the burden is the least restrictive means of achieving that compelling interest.[40]

The third element, the "compelling governmental interest" clause, must be read to account for the experience and expertise of prison and jail administrators in establishing necessary regulations and procedures to maintain good order, security, and discipline, consistent with consideration of costs and limited resources.[41] This means that RLUIPA is not meant to elevate accommodation of religious observances over an institution's need to maintain order and safety.[42]

In the fourth element, "least restrictive means" is a relative term that implies a comparison with other means.[43] This requires the government to "acknowledge and give some consideration to less restrictive alternatives" to determine whether an alternative might be equally as successful as the policy in furthering the identified compelling interests.[44] In carrying this burden, the government does not need to conceive of and then reject every possible alternative.[45] Rather, it must demonstrate that it considered and rejected the other options brought to the government's attention.[46]

D. Harms of Deferring Discretion to the "Resources, Experience, and Expertise of Prison and Jail Administrators" in Private Prison Business Models

The U.S. prison population experienced a significant spike between 1970 and 1990, leaving states to manage prison overcrowding with limited funds.[47] State spending on prisons today is

[40] *Id.*
[41] *Cutter,* 544 U.S. at 723 (2005).
[42] *Id.*
[43] *Faver v. Clarke*, No. 19-7634, 2022 U.S. App. LEXIS 2934 (4th Cir. 2022).
[44] *Id.*
[45] *Id.*
[46] *Id.*
[47] Gaby Galvin, *Underfunded, Overcrowded State Prisons Struggle With Reform*, U.S. NEWS AND WORLD REPORT, Jul. 26, 2017, https://www.usnews.com/news/best-states/articles/2017-07-26/understaffed-and-overcrowded-state-prisons-crippled-by-budget-constraints-bad-leadership.

influenced by a range of factors, and costs can vary widely from year to year.[48] Some factors are primarily outside of a corrections department's control, such as rising health care costs, statewide policies to increase public employee salaries and benefits, and policy changes that influence sentencing, thus affecting the size of the prison population.[49] Other factors are more within the corrections department's control, such as inmate-to-officer staffing and the number of prison facilities the state operates.[50]

Many states reduce their prison spending, despite a growing number of people in state prisons.[51] These states run the risk of lowering the office-to-inmate staffing ratio available to inmates.[52] This increases the potential threats to both staff and incarcerated people.[53] A decline in spending despite a growing prison population also decreases the state's ability to offer services, programs, or treatment, resulting in worse outcomes, including leaving mental, behavioral, and other health needs unaddressed and increasing recidivism if people are not adequately prepared for reentry.[54]

Private prison budgeting decisions play an instrumental role in the resources and flexibility it awards to its prisoners. Suppose a correctional facility has plentiful resources, therefore having adequate staffing and secure premises. In that case, it is likely to invest the time and effort to accommodate their prisoners' religious needs. Unfortunately, this conflicts with the private-prison business model that often seeks to identify ways to cut costs, at the expense of prison resources, in an effort to maximize profit. Today, most prisons are dangerously understaffed, as one-quarter of

[48] Chris Mai & Ram Subramanian, *Price of Prisons: Examining State Spending Trends, 2010-2015,* VERA INSTITUTE OF JUSTICE, 21, May 2017, https://storage.googleapis.com/vera-web-assets/downloads/Publications/price-of-prisons-2015-state-spending-trends/legacy_downloads/the-price-of-prisons-2015-state-spending-trends.pdf.
[49] *Id.*
[50] *Id.*
[51] *Id*.
[52] *Id*.
[53] *Id*.
[54] Mai & Subramanian, *supra* note 43.

all correctional officer jobs in the United States are vacant.[55] This is likely due to the low pay and high turnover with officers lured away by the state police or local sheriffs that pay more.[56] This raises serious concerns as correctional officers are often not adequately trained and inexperienced.[57]

The issue then becomes how prisons and their staff possess the requisite resources to "maintain good order, security, and discipline, consistent with consideration of costs and limited resources" while accommodating their inmates' religious needs.[58] The court in *Cutter* has failed to recognize that although it is reasonable to emphasize the need for relative accommodation toward religious practices, many private prisons either do not have the adequate resources to adequately staff their facilities or even provide basic hygiene supplies to their inmates.[59] This is especially true when the prison is profitable and can cut extra funding for religious practices while still "maintain[ing] good order, security, and discipline, consistent with consideration of costs and limited resources," as required today.

This reality poses a severe threat to the exercise of all religions within a prison. Two Muslim inmates sued prison officials in California, saying they were forced to eat food forbidden by their faith.[60] In Mississippi, Christian inmates sued, claiming that prison officials violated their First Amendment rights by refusing to allow inmate-led services and prohibiting inmates from

[55] Sandy Hausman, *State Prisons Are Dangerously Understaffed*, RADIO IQ WYTF, Sep. 7, 2021, https://www.wvtf.org/news/2021-09-07/state-prisons-are-dangerously-understaffed.
[56] *Id*.
[57] *Id*.
[58] *Cutter v. Wilkinson,* 544 U.S. 709, 723 (2005).
[59] *Smith v. Gomez*, No. 95-15773, 1996 U.S. App. LEXIS 11611, *2 (9th Cir. May 3, 1996) (discussing an instance in which an inmate had to choose between paying photocopying costs and purchasing hygiene supplies. The prison officials submitted evidence that basic personal hygiene supplies were provided at state expense and that the budget was too low to afford both for all inmates).
[60] David L. Hudson, *Prisoners Rights*, FREEDOM FORUM INSTITUTE, 2008, https://www.freedomforuminstitute.org/first-amendment-center/topics/freedom-of-religion/free-exercise-clause-overview/prisoners-rights/.

preaching.[61] A Jewish inmate in Ohio also sued prison officials after cutting his beard, which he says was necessary for his faith.[62]

It then becomes evident that due to the lack of appropriate resources, staffing, and training, correctional institutions in the United States are becoming less reliable in their abilities to accommodate inmates' religious needs.[63] To compound this harm, prison staff along with their governing authorities are protected from any liability so long as the restrictive prison policy accounts for the "experience and expertise of prison and jail administrators in establishing necessary regulations and procedures to maintain good order, security, and discipline, consistent with consideration of costs and limited resources."[64]

This standard fundamentally fails to recognize that the inexperience of prison officers, and lack of resources available to them, is not the responsibility of the inmates, but of the state. Additionally, the private prison business model naturally incentivizes prisons to provide the minimum to their inmates in an attempt to profit the rest. An inmate has very little control over where they will be incarcerated or what resources will be available to them once they arrive. This means that the legitimacy of their First Amendment free exercise claim is evaluated given the subjective standard of prison resources, leadership, and discretion, not any objective standards of uniformity. This evaluation can cause inconsistent results, and often obstructs an inmate from freely practicing their respective religion.

E. What Can be Done

[61] *Id.*; *see also O'Lone v. Estate of Shabazz*, 482 U.S. 342, (1987) (where a group of Muslim inmates challenged New Jersey state prison policies that prohibited them from attending Jum'ah, a weekly Muslim religious ceremony. Prison officials determined that allowing inmates to leave outside work details to go inside the prison for the religious service would imperil the safety and institutional order. They testified that inmates returning from outside work details created too much congestion and delays at the main gate, which is a high-risk area).
[62] *Id.*
[63] Hausman, *supra* note 55.
[64] *Cutter v. Wilkinson,* 544 U.S. 709, 723 (2005).

The current evaluation of a prisoner's right to exercise their religion fails to take into consideration any objective standards pertaining to prisons providing adequate resources to their inmates. By putting the burden of proof on the state to prove that the correctional officers acted within their experience and expertise and established the necessary regulations and procedures to maintain good order, security, and discipline, consistent with consideration of costs and limited resources, the court fails to evaluate who the actors of the correctional system are and what resources they have.[65]

The court must make a more objective, reasonable person approach to Free Exercise and RLUIPA claims. Courts or legislatures should put the burden on the state to provide reasonable accommodations to its inmates regardless of the resources allocated to the prison or the training given to the officers. By establishing this bright line, objective rule, state officials will be compelled to provide adequate resources to their prisons so that all inmates have a reasonable, uniform expectation of religious freedom regardless of where they are incarcerated. This rule would have significantly beneficial implications for both the inmates and society at large.

F. Why Reform is Necessary

One benefit of holding prisons to an objective and relatively high standard of care is that it will improve prison conditions.[66] This is because religion often helps inmates deal with the emotional strains of incarceration.[67] Religion not only explains the cause of one's failure but also prescribes the solution.[68] There are three general ways in which prisoners express how religion can ease their pains of incarceration.[69]

---

[65] *Id.*
[66] *See* Clear, et al., *supra* note 8 at 7.
[67] *Id.* at 4.
[68] *Id.*
[69] *Id.*

First is the notion of dealing with guilt.[70] Apart from the loss of freedom, the most powerful message of imprisonment is guilt.[71] Prisoners can turn to religion for relief, as a kind of exculpatory acceptance of the workings of evil in the world, or as atonement and forgiveness.[72] Clinical theory, clinical observations, and qualitative and quantitative empirical studies indicate that when people feel guilty about a specific behavior, they are motivated to confess, apologize, and repair their wrongdoings.[73] They are also more empathic with the victims of their misdeeds and more likely to take responsibility, and make amends.[74] Allowing prisoners to freely exercise their religion can facilitate legitimate opportunities for inmates to exemplify guilt and remorse.

The second is that religion helps find a new way of life for prisoners.[75] One of the main themes religious inmates provided in discussing the importance of their faith is that it "changed" them.[76] The vast majority of data suggests the importance of religious influences in protecting individuals from harmful outcomes as well as promoting beneficial outcomes.[77] The beneficial relationship between religion and crime reduction is not simply a product of religion's constraining function or what it discourages (e.g., opposing drug use or delinquent behavior), but also a matter of what it encourages (e.g., promoting positive behaviors).[78]

---

[70] *Id.*
[71] *Id.*
[72] *Id.*; *See also* June P. Tangney et al., *Assessing Jail Inmates' Proneness to Shame and Guilt: Feeling Bad About the Behavior or the Self?*, 38(7) CRIM. JUSTICE BEHAV. 710 (2011) ("Shame and guilt are emotions with special relevance to the field of criminology, and to the rehabilitation process. They are generally regarded as quintessential 'moral' emotions because of their presumed roles in inhibiting immoral, socially undesirable behavior and in fostering altruistic, prosocial behavior.").
[73] Tangney, et al., *supra* note 72 at 711.
[74] *Id.*
[75] Clear, et al., *supra* note 8.
[76] *Id.*
[77] Byron R. Johnson & Curtis S. Schroeder, *Religion, Crime, and Criminal Justice*, OXFORD UNIVERSITY PRESS, 2014, http://www.baylorisr.org/wp-content/uploads/ReligionCrime-and-Criminal-Justice.pdf.
[78] *Id.* at 9-10.

The third way that religion can ease the pains of incarceration is by dealing with the loss of freedom.[79] The combination of the deprivations imposed by prison life and the negativistic culture imported into it make the prison environment a difficult setting in which to live.[80] Religious programs can be designed to ameliorate the harsh environment of the prison such as the lack of safety, material comforts, and sexual contact.[81]

In addition to improving the quality of life for inmates within a prison, religious opportunities are an evidence-based method in providing a prosocial behavioral pattern that is proven to combat recidivism.[82] Prosocial behaviors are those intended to help other people and are characterized by a concern for the rights, feelings, and welfare of others.[83] Behaviors that can be described as prosocial include not just feeling empathy and concern for others, but actually behaving in ways that help or benefit other people.[84] The freedom to participate in religious practices and activities can foster the development of and integration into personal networks that provide both social and emotional support.[85]

When such personal networks overlap with other networks, they not only constrain illegal behavior but can also protect one from the effects of living in disadvantaged communities.[86] This indicates that, an individual's integration into a community-based religious network weakens the

[79] Clear, et al., *supra* note 8.
[80] *Id.* at 6.
[81] *Id.*
[82] Byron R. Johnson, *How Religious Freedom Contributes to Positive Criminology and Justice Reform*, RELIGIOUS FREEDOM INSTITUTE, (Dec. 2020), https://www.religiousfreedominstitute.org/cornerstone/how-religious-freedom-contributes-to-positive-criminology-and-justice-reform.
[83] *Id.*
[84] *Id.*
[85] *Id.*; *see also, SpearIt, Religion as Rehabilitation? Reflections on Islam in the Correctional Setting,* 34 WHITTIER L. REV. 29, 34 (2012) ("The Department of Health and Human Services reports that literature on religion and recidivism 'is consistent with criminological theories supporting the claim that religious beliefs are inversely related to delinquency, crime and recidivism.' In the juvenile context, some researchers assert that high levels of participation in religious activities can reduce juvenile delinquency").
[86] *Id.*

effects of other factors that might otherwise promote deviant behavior.[87] Thus, religious networks have been found to buffer or shield people from the effects of harmful influences. [88]

The final way in which improving religious exercise freedoms in prison would be beneficial is that it would mitigate the risk of negative self-labeling. Generally, human behavior is a product of social interaction and conversation.[89] Researchers have since examined self-fulfilling prophecies and self-identification in a variety of persons.[90] Research indicates that all social groups formulate rules, laws, or guidelines and attempt to enforce them.[91]

Ordinarily, once a rule is enforced, the individual who has broken it may be seen as a unique kind of person who cannot be trusted to live by the rules agreed on by the group, causing them to be regarded as an outsider.[92] The individual labeled as an outsider may have a different view of the matter and may not accept the rule by which he or she is being judged.[93] The person may regard those who cast judgment on them as either incompetent or illegitimately entitled to do so.[94] Negative effects of poor labeling cause an individual to be isolated and eventually ostracized by mainstream groups.[95] This marginalization leads to a clustered self-perception that includes a negative self-image that is perpetuated by society to develop relationships and associations with each other that lead to a higher likelihood of criminal activity.[96]

---

[87] *Id.*
[88] *Id.*
[89] Ross L. Matsueda, *LABELING THEORY*, 13-44 (David P. Farrington & Joseph Murray, 1st ed. 2017).
[90] Howard Becker, OUTSIDERS – DEFINING DEVIANCE, 1 (THE FREE PRESS, 1963), https://leeclarke.com/courses/intro/readings/becker_definingdeviance.pdf.
[91] *Id.*
[92] *Id*.
[93] *Id*.; *see also* W. David Ball, *The Civil Case at the Heart of Criminal Procedure: In re Winship, Stigma, and the Civil-Criminal Disctinction,* 38 AM. J. CRIM. L. 117, 146, 2011 (noting that "one initial theory posited that stigmatic labels in some way acted as self-fulfilling prophecies. This theory, "labeling theory," suggested that individuals who were labeled as deviants internalized those labels and became the deviants they were branded to be").
[94] *Id*.
[95] *Id*.
[96] *Id.*

There are three types of labels: self-defining, ascribed, and achieved.[97] Self-defining means you choose what label you want for yourself.[98] The prime example of this is religion. If one can self-define themselves as a follower of a given religion and diverge from the label of criminal deviant, this is likely to be advantageous when assessing prison behavior and overall recidivism.

G. Data on Religion and Recidivism

Data suggests that religious exposure (both during and after incarceration) is an evidence-based method in combatting future crime.[99] When religion is associated with reduced criminal offending, it is often because faith-based interventions in an institutional setting began or accelerated the religious or spiritual journey to desistance.[100] It is often the case that an individual does not have his or her exposure to organized religion until they are incarcerated, which is the initial starting point in deterring recidivism. With religious opportunity, people metaphorically subscribe toward a pro-social self and are connected to pro-social peers and institutions.[101] Then, continued strengthening and support develop a socialized mind and integrate individuals into the larger social context.[102] When this happens, pro-social relationships and opportunities can be built thereafter, but structural changes in bonds and opportunities are unlikely to lead to desistance

[97] Johannes Knutsson, *Labeling Theory: A Critical Examination,* SCIENTIFIC REFERENCE GROUP, (Mar. 1977), https://www.ojp.gov/pdffiles1/Digitization/-47664NCJRS.pdf.
[98] *Id.*
[99] *See* Thomas J. Mowen, et al., During, *After, or Both? Isolating the Effect of Religious Support on Recidivism During Reentry*, 34(4) J Quant Criminol. 1079–1101 (2018) "For criminal offending, results demonstrate that baseline levels of religious support and within-individual changes in religious support do not, by themselves, relate to offending post-release. However, in complicating this, findings reveal that it is the joint combination of the two that protect against offending, even while controlling for offending prior to incarceration. This is an important methodological step in disentangling the role of religious support during reentry as our findings demonstrate that ongoing support post-incarceration feeds off baseline religious support to reduce the likelihood of criminal reoffending. Methodologically, this finding demonstrates that failing to account for how baseline differences in religious support condition the relationship of religious support across time on offending would overlook the importance of religious support in both capacities."
[100] *Id.*
[101] Ray Paternoster & Shawn Bushway, *Desistance and the "feared self": toward an identity theory of criminal desistance.* 99 J CRIM. LAW AND CRIMINOLOGY 1103, 1103–1156 (2009).
[101] *Id.*
[102] *Id.*

without prior personal change or initial contact with religion in prisons.[103] Studies illustrate that the combination of initial and post-release support facilitates an ability to deal with the uncertain and complex demands of life and fuels a criminal's transformation away from recidivism.[104]

Another key issue in deterring recidivism is dealing with substance abuse. In the United States, roughly 85% of the prison population has an active substance use disorder or were incarcerated for a crime involving drugs or drug use.[105] Studies show that 73% of addiction treatment programs in the nation include a spirituality-based element and the vast majority of which emphasize reliance on God or a Higher Power to stay sober.[106] Additionally, more than 84% of scientific studies show that faith is a positive factor in addiction prevention or recovery.[107] Many scientists indicate that decline in religious affiliation in the United States is not only a concern for religious organizations but constitutes a national health concern when it comes to the treatment and ultimate recovery of substance abusers in prisons.[108]

H. The Importance of Change in Jurisprudence or Law

Understanding the trends and tendencies of judicial decision-making is often a complex and multilevel analysis. In a study conducted in 2015, researchers sought to understand judicial decision-making dynamics by using a sample of 330 federal cases from 2000 – 2007 to examine the legal and nonlegal factors that are suspected to affect Free Exercise claims.[109] It was found that

[103] Mowen, *supra* note 11.
[104] R Kegan, *In Over Our Heads* (1994).
[105] National Institute on Drug Abuse, *Criminal Justice DrugFacts*, https://nida.nih.gov/publications/drugfacts/criminal-justice#:~:text=85%25%20of%20the%20prison%20population,involving%20drugs%20or%20drug%20use.
[106] Brian J. Grim & Melissa Grim, *Belief, Behavior, and Belonging: How Faith is Indispensable in Preventing and Recovering from Substance Abuse*, 58 J Relig Health, 1713, 1715 (2019).
[107] *Id*.
[108] *Id*.
[109] Benjamin Meade & John D. Burrow, *Untangling the Dynamics of Judicial Decision Making and Inmates' Free Exercise Claims*, 95 THE PRISON JOURNAL 3, 16, (2015).

legal rules, more than any other variable tested, were among the strongest predictors of case outcomes.[110]

The findings suggest that the law itself exerts a powerful effect on the ultimate decision to grant or deny a request for a religious accommodation.[111] Notwithstanding the merits of RFRA as precedent, judges still possess significant latitude when deciding what are the appropriate legal principles to use to dispose of Free Exercise claims.[112] This magnifies the need for a comprehensive reevaluation on how the Free Exercise Clause and RLUIPA are interpreted through jurisprudence transformation or new law rather than any other form of advocacy.[113]

III. Conclusion

The First Amendment guarantees two different types of religious freedom.[114] The Establishment Clause ensures that the government cannot advance a specific religion, or heavily restrict religion, while the Free Exercise Clause seeks to protect people's religious beliefs and practices.[115] The First Amendment is one of the most important amendments for the protection of democracy.[116] Freedom of religion allows people to believe and practice whatever religion they want.[117]

---

[110] *Id.*
[111] *Id.*
[112] *Id.*
[113] *See* 1 Federal Standards of Review § 6.01 (2021) ("A district court is bound by case law from the appellate court within its circuit. Some courts, however, say the interests of justice are not served by rigid adherence to precedent that is clearly and substantially undercut by later circuit decisions. The decisions of the district courts themselves are not binding on fellow district judges or on the circuit courts and Supreme Court, except perhaps in the same line of litigation under the doctrine of law of the case. The district court's opinion in any case may be considered for its reasoning and support, and some courts have given 'great weight' to certain trial court decisions, even outside the circuit").
[114] U.S. CONST. amend. I.
[115] *Id.*
[116] *What Is the First Amendment and What Does It Do?*, VOA NEWS, Nov. 2018, https://www.voanews.com/a/whatisthefirstamendmentandwhatdoesitdo/4662700.html#:~:text=The%20First%20Amendment%20is%20one,without%20the%20government%20stopping%20them.
[117] *Id.*

The current approach to evaluating free exercise claims presents a highly problematic and ununiformed view of our flawed correctional system. The approach fails to consider that many correctional facilities operate with differing access to financial resources.[118] This means that individuals incarcerated in a more financially stable facility will have access to more religious freedoms than those serving their prison sentences in poorer facilities. This is because prisons with less financial resources often have a poor officer-to-inmate ratio and struggle to maintain the requisite safety within their walls.

Unfortunately, most prisons in the United States have scarce financial resources and few employed qualified correctional officers.[119] This means that courts are more likely to favor a correctional institution in Free Exercise claims even when the plaintiff can show that their religious exercise has been burdened and that the burden is substantial. This standard also assumes that institutions do everything in their power to help inmates when that should not be the assumption.

The inability to practice one's religion in prison has a variety of collateral consequences. First, it is inhumane to force two Muslim inmates to eat food forbidden by their faith.[120] It's also immoral to prohibit Christians from having an inmate-led service and outlawing preaching and forcing a Jewish man to cut his beard, which is necessary for his faith.[121] Second, allowing inmates to practice their religion freely can improve prison conditions by enabling inmates to deal with guilt, find a new way of life, and ease the pains of incarceration by coping with the loss of freedom.[122] Third, an individual's integration into a community-based religious network weakens the effects of other factors that might otherwise promote deviant behavior.[123] Finally, religion is a

[118] Galvin, *supra* note 87.
[119] *Id.*
[120] Galvin, *supra* note 87.
[121] *Id.*
[122] Clear, et al., *supra* note 8.
[123] SpearIt, *supra* note 85.

positive way to mitigate the risk of negative self-labeling and an evidence-based method in combatting recidivism.[124] Rather than any other form of advocacy, a comprehensive change in jurisprudence is essential in this change as judges are likely to favor black-letter law over any other variable in Free Exercise cases.[125]

With the United States facing its most imminent and serious crime wave in centuries, legislatures, bureaucrats, and even the judiciary have taken a keen interest in mitigating the possibility of increased violence.[126] One way to accomplish this goal is to replace the current approach to incarcerated persons' Free Exercise claim. It would be advantageous for the courts or the legislature to adopt an objective, reasonable person standard that is not influenced by financial resources. With government contracts, private prisons have a fundamental obligation to finance their facilities adequately. The private prison business model, however, is likely to cut any extra resources provided in an effort to maximize profit. If the court or legislature were to implement an objective, reasonable person standard, all inmates, regardless of where they are incarcerated, will have the uniform opportunity to practice their religions freely. Identifying what practices are guaranteed within the bright-line rule should be left to the courts or legislature to decide.

[124] Matsueda, *supra* note 89.
[125] Benjamin & Burrow, *supra* note 109.
[126] The United States Drug Enforcement Association, *Part One: DEA-Washington is Standing Up Against Violent Crime*, ("Homicide rates and violent crimes [are] increasing to a disturbing high"), Jan. 20, 2022, https://www.dea.gov/stories/2022/2022-01/2022-01-20/part-one-dea-washington-standing-against-violent-crime.

QUALIFIED OPPORTUNITY ZONES: A CRITICAL TAX ANALYSIS OF THE CAPITAL GAIN REQUIREMENT

By Islame Hosny*

In December 2017, the Tax Cuts and Jobs Act established the Qualified Opportunity Zones (QOZ) tax incentive program, "an economic development tool that allows people to invest in distressed areas."[1] As a student at Columbia University, which has its main campus in the affluent Morningside Heights neighborhood in New York City, I sometimes pass through Central Harlem as I drive north on Amsterdam Avenue to make my way home to New Jersey. The moment I pass between 122nd Street and 123rd Street, the point where Morningside Heights ends and Central Harlem begins, I notice a striking difference between the two neighborhoods: one is luxurious and wealthy, while the other is stricken with poverty. Because of its economic plight, Central Harlem is designated as a QOZ.

The purpose of the QOZ program is to "spur economic development and job creation in distressed communities by providing tax benefits to investors."[2] The program encourages investment by those who have capital gains into economically troubled neighborhoods. As a tax attorney, I sometimes help clients understand the mechanics of the requirements they need to satisfy to obtain the tax benefits of investing in Qualified Opportunity Funds (QOFs), which are investment vehicles used to invest in QOZs. But the juxtaposition of these two adjacent – but remarkably different – neighborhoods piqued my curiosity to know more about how the QOZ

* J.D., Fordham University School of Law, LL.M., New York University School of Law. Islame is an associate attorney at the law firm of Pryor Cashman LLP and is a student at the Columbia University School of General Studies.

[1] *Opportunity Zones*, INTERNAL REVENUE SERV., https://www.irs.gov/newsroom/opportunity-zones (Apr. 2022).

[2] *Id*.

program impacts QOZ residents, not just QOF investors. One important aspect of the program is the capital gain requirement: an investor can only receive the tax benefits of the QOZ program with respect to capital gains they invest in a QOF. Because 29%[3] of Central Harlem residents are below the federal poverty level, and since most capital gains are realized by high-income families,[4] Central Harlem residents are unlikely to have any capital gains to invest. Thus, unlike affluent investors who are more likely to have capital gains, Central Harlem residents might not be able to receive any of the tax benefits of the QOZ program. This raises an important question: from a tax equity perspective, is the QOZ program fair to both QOF investors and QOZ residents?

One of the touchstones of traditional tax policy is that fairness is an essential consideration in determining who should bear a tax. Fairness of a tax has been discussed in traditional tax policy literature as having two components: vertical equity and horizontal equity.[5] Horizontal equity suggests that similarly situated taxpayers should be treated similarly.[6] Meaning, those with equal ability to pay should pay equal taxes.[7] Vertical equity indicates that those who earn more should be taxed more than those who earn less.[8] In the U.S., we have a progressive tax system, which means that as a person's income increases, so does the tax rate on their income – consistent with the principle of vertical equity.

[3] *Central Harlem Manhattan: Commercial District Needs Assessment*, N.Y.C. DEP'T OF SMALL BUS. SERVS. 17, https://www1.nyc.gov/assets/sbs/downloads/pdf/neighborhoods/avenyc-cdna-centralharlem.pdf (last visited May 6, 2022).

[4] Carl Davis et al., *State Income Taxes and Racial Equity: Narrowing Racial Income and Wealth Gaps with State Personal Income Taxes*, INST. ON TAX'N & ECON. POL'Y 16 (Oct. 2021), https://itep.sfo2.digitaloceanspaces.com/State-Income-Taxes-and-Racial-Equity_ITEP_October2021.pdf.

[5] LAURIE L. MALMAN ET AL., THE INDIVIDUAL TAX BASE: CASES, PROBLEMS AND POLICIES IN FEDERAL TAXATION 8, 14 (2d ed. 2002).

[6] Nancy, J. Knauer, *Critical Tax Policy: A Pathway to Reform?,* 9 NW. J. L. & SOC. POL'Y, 206, 222 (2014).

[7] C. EUGENE STEUERLE, *Contemporary U.S. Tax Policy* 10, URBAN INSTITUTE PRESS (2004).

[8] Knauer, *supra* note 6, at 222.

The concept of progressive tax rates is fundamentally rooted in the diminishing marginal utility of income theory which suggests that the utility a person derives from each additional dollar of income is less than the utility derived from each previous dollar.[9] The reasoning behind the theory is that earlier dollars are used first for necessities such as food, shelter, clothing, and healthcare. The diminishing marginal utility of income theory reinforces our progressive tax system by providing justification for imposing higher tax rates on those who have more income and have the wherewithal to pay more tax. A progressive tax system helps reduce the wealth gap by redistributing wealth from people who have higher income to those who have lower income, which in turn helps reduce racial disparities.[10] Thus, to analyze the fairness of the QOZ program, the starting point should be understanding whether the program is consistent with our progressive tax system.

The QOZ program offers three tax incentives. The first incentive is deferral, until December 31, 2026, of federal income tax on capital gains that are invested in a QOF.[11] The second incentive is a tax basis step-up, which effectively results in an exclusion from federal income taxation of up to 15% of the deferred gain.[12] The third incentive occurs when an investor holds its interest in a QOF for at least 10 years[13] and sells the investment on or before December 31, 2047.[14] If the 10-year holding period requirement is satisfied, the investor's tax basis in their interest in the QOF is adjusted to its fair market value – effectively resulting in a 100% exclusion from federal

[9] *See* Mark Stein, *Diminishing Marginal Utility of Income and Progressive Taxation: A Critique of The Uneasy Case*, 12 N. ILL. U. L. REV. 373, 374 (2004).

[10] Chye-Ching Huang & Rodrick Taylor, *How the Federal Tax Code Can Better Advance Racial Equity: 2017 Tax Law Took a Step Backward*, CTR. ON BUDGET & POL'Y PRIORITIES 2 (Jul. 25, 2019), https://www.cbpp.org/sites/default/files/atoms/files/7-25-19tax.pdf.

[11] I.R.C. § 1400Z-2(b)(1) (2017).

[12] *Id.* § 1400Z-2(b)(2).

[13] *Id.* § 1400Z-2(c).

[14] Treas. Reg. § 1.1400Z2(c)-1(c) (as amended in 2020).

income taxation on appreciation on the QOF investment.[15] Compared to the first two incentives, the third has the potential to provide the greatest benefit to a QOF investor: a 0% effective federal income tax rate on a potentially unlimited amount of gain. But how does that effective tax rate compare to that of a QOZ resident?

In Central Harlem, the median household income is $55,870.[16] Because an estimated 86.3% of taxpayers take the standard deduction,[17] a Central Harlem resident is less likely to itemize their deductions and more likely to take the standard deduction. While the average household in Central Harlem consists of approximately two people,[18] 42.7% of Central Harlem households are single-person households, 77.2% of Central Harlem households do not include children under age 18,[19] and about 70% of Central Harlem Residents are not married.[20] Therefore, a typical Central Harlem resident would have a tax filing status of "single" and would take the standard deduction.

The federal income tax liability for a single filer with income of $55,870 who takes the standard deduction ($12,550)[21] and, thus, has taxable income of $43,320 is $5,280[22] – resulting in an effective federal income tax rate of 12.19%. A single filer with no qualifying children and who

[15] *See* Dana McCartney et. al., *Tax Advantages of QOZ Investments*, THE TAX ADVISER (Dec. 1, 2019), *available at* https://www.thetaxadviser.com/issues/2019/dec/tax-advantages-qoz-investments.html.

[16] *Central Harlem MN10*, FURMAN CTR. FOR REAL ESTATE & URB. POL'Y, N.Y.U., https://furmancenter.org/neighborhoods/view/central-harlem (last visited May 6, 2022).

[17] Scott Eastman, *How Many Taxpayers Itemize Under Current Law?*, TAX FOUND. (Sep. 12, 2019), https://taxfoundation.org/standard-deduction-itemized-deductions-current-law-2019/.

[18] *Central Harlem Manhattan: Commercial District Needs Assessment*, *supra* note 3, at 16.

[19] *Central Harlem MN10*, *supra* note 16.

[20] *NYC-Manhattan Community District 10--Central Harlem PUMA, NY*, CENSUS REPORTER, https://censusreporter.org/profiles/79500US3603803-nyc-manhattan-community-district-10-central-harlem-puma-ny/ (last visited May 6, 2022).

[21] 1040 (and 1040-SR) Instructions (Tax Year 2021), INTERNAL REVENUE SERV., 6 (Dec. 21, 2021), https://www.irs.gov/pub/irs-prior/i1040gi--2021.pdf.

[22] *Id.* at 70.

has adjusted gross income in the amount of $43,320 is not eligible for the child tax credit[23] or the earned income tax credit.[24] Thus, a typical Central Harlem resident would not be able to benefit from the earned income tax credit or the child tax credit to reduce their effective federal income tax rate below 12.19%. As a result, a typical Central Harlem resident would have a higher effective federal income tax rate than a QOF investor.

When a QOF investor's effective federal income tax rate is compared to that of a typical Central Harlem resident, the QOZ program seems to be inconsistent with our progressive tax system. In fact, the QOZ program has a regressive effect by taxing people with higher income (QOF investors) at lower tax rates than people with lower income (QOZ residents). But, although traditional tax equity analysis highlights the regressive impact of the QOZ program, it does not shed light on whether the program disproportionately impacts any specific demographic groups. To answer that question, it may be helpful to analyze the QOZ program through the lens of critical tax theory, a growing body of scholarship seeking to uncover explicit and implicit bias in the tax system.

Critical tax theorists use various methods of inquiry, including the interpretation of "social science and economic data to show how the tax law impacts groups differently."[25] Critical tax theory is not "the only way (or the best way) to approach tax law;" it is merely "a lens that one can pick up and put down."[26] While traditional tax policy can be theoretical and tends to focus on

[23] *See* INTERNAL REVENUE SERV., PUBLICATION 972, CHILD TAX CREDIT AND CREDIT FOR OTHER DEPENDENTS (For use in preparing 2020 Returns) 3 (Jan. 11, 2021), *available at* https://www.irs.gov/pub/irs-prior/p972--2020.pdf.

[24] *See* INTERNAL REVENUE SERV., *Earned Income and Earned Income Tax Credit (EITC) Tables*, https://www.irs.gov/credits-deductions/individuals/earned-income-tax-credit/earned-income-and-earned-income-tax-credit-eitc-tables (last visited May 6, 2022).

[25] Anthony C. Infanti & Bridget J. Crawford, Critical Tax Theory: An Introduction xxvi-xxvii (2009).

[26] *Id.* at xxviii.

income-related attributes of taxpayers, critical tax theory can bolster traditional tax policy by analyzing the practical consequences, both intended and unintended, of the tax system on certain subsets of taxpayers. In that way, critical tax theory does not replace, but can enhance, traditional tax policy by adding an additional dimension to tax equity analysis. In applying a critical tax lens, however, it is important to understand one fundamental difference between traditional tax policy and critical tax theory, which lies in how each school of thought views the presumption of taxpayer neutrality.

Embedded in traditional tax policy is a "strong presumption of taxpayer neutrality where the only salient distinction among taxpayers is that of income level."[27] Proponents of critical tax theory argue that relying on income as the sole metric for analyzing tax equity assumes that people who have the same amount of income are impacted the same way by the tax system. Additionally, critical tax theorists argue that by not considering certain taxpayer attributes, such as race, ethnicity, sexual orientation, disability status, gender identity, or socioeconomic status, the tax system fails to consider any disproportionate impact the tax system may have on members of various demographic groups.[28] To apply a critical tax lens, enhanced information collection on taxpayer demographics is crucial because it would permit policymakers to understand any disparate impact the tax system has on different demographic groups.[29] Critics of critical tax theory argue that those who apply a critical tax lens have no reliable data to support their claims.[30] However, that claim incorrectly assumes that the lack of data showing a disproportionate impact of the tax system on various demographic groups equates to proof that the tax system is equitable.

[27] Knauer, *supra* note 6, at 209.
[28] *See id.* at 209-11.
[29] *See id.* at 247.
[30] *See* Richard Schmalbeck, *Race and the Federal Income Tax: Has a Disparate Impact Case Been Made?,* 76 N.C. L. REV. 1817, 1823 (1998).

The lack of reliable demographic data for taxpayers highlights the need for data collection such that those data can be analyzed to determine whether taxpayer income level is the only relevant factor in tax equity analysis.

Yet, this need for taxpayer demographic information is in tension with one of the pillars of the U.S. tax system – the confidentiality of taxpayer information. I.R.C. § 6103 prohibits the disclosure by the government of any taxpayer information.[31] While taxpayer demographic data is needed to perform critical tax equity analysis, divulging that information to the public for analysis would violate taxpayer confidentiality. Adding to the dilemma is the fact that the IRS does not collect taxpayer demographic information such as race and ethnicity.[32] The lack of data collection on taxpayer demographics by the IRS combined with strict prohibitions on disclosure of taxpayer information would seem to create an impediment for critical tax equity analysis. However, the Treasury Department has proposed a potential solution to both predicaments: the creation of reliable synthetic data.

On December 14, 2021, the Treasury Department confirmed its commitment to advancing equity analysis in tax policy by "examining the tax system through a racial equity lens."[33] As part of this commitment, the Treasury Department has proposed to work with other federal agencies and data production partners to develop a reliable methodology to produce synthetic data by utilizing statistical modelling techniques to impute race, ethnicity, gender, and other demographic characteristics for tax data. The resulting synthetically modelled data would then be utilized for

[31] I.R.C. § 6103 (2020).

[32] *See* Wally Adeyemo & Lily Batchelder, *Advancing Equity Analysis in Tax Policy,* U.S. DEP'T OF THE TREASURY (Dec. 14, 2021), https://home.treasury.gov/news/featured-stories/advancing-equity-analysis-in-tax-policy.

[33] *See id.*

tax equity analysis. Thus, synthetic data could resolve the conflict between the need for more demographic taxpayer data and the requirement to maintain confidentiality of taxpayer information – paving the way for meaningful critical tax equity analysis on a national scale. Even though the government has not yet released these synthetic data to the public, we can perform a preliminary critical tax equity analysis to approximate the impact of the QOZ program on Central Harlem residents by relying on census data and other publicly available information.

To apply a critical tax lens to the QOZ program, we turn back to the capital gain requirement. As briefly mentioned earlier, the tax benefits of the QOZ program are limited to those who invest their capital gains in QOFs. However, most capital gains are realized by high-income households.[34] In 2018, 85% of capital gains were reported by households with adjusted gross income of more than $200,000 and only 7% of capital gains are received by households with adjusted gross income under $100,000.[35] "Because capital gains overwhelmingly flow to high-income households, they also tend to flow mostly to white households."[36] The median household income in Central Harlem, $55,870,[37] is far below the $200,000 figure mentioned above. Additionally, 77.9% of Central Harlem residents are Black or Hispanic and only 15.5% are white.[38] Based on these data, it is unlikely that Black or Hispanic residents of Central Harlem would have any capital gains to invest, and it is much more likely that white taxpayers invest in the Central Harlem QOZ. Exacerbating the problem is the fact that most QOF investors are not likely to reside in QOZs.[39] Therefore, the capital gain requirement has the practical effect of creating a structural

[34] *See* Davis et al., *supra* note 4, at 16.
[35] *Id.*
[36] *Id.*
[37] *See supra* note 16 and accompanying text.
[38] *Central Harlem MN10*, *supra* note 16.
[39] Monica L. Wendel & Gabriel Jones Jr., *Equity for Whom? The Example of Qualified Opportunity Zones*, 110 AM. J. PUB. HEALTH, 280, 281 (2020).

barrier that is likely to exclude Black and Hispanic residents of Central Harlem from the tax benefits of the QOZ program, while limiting those benefits to outside investors who are most likely white. Such a one-sided allocation of tax benefits is likely to increase the existing racial wealth gap.

The median Black household and the median Hispanic household have 13 cents and 16 cents, respectively, in wealth for each dollar in wealth held by a median white household.[40] The 0% and 12.19% effective federal income tax rates that apply to a QOF investor and typical Central Harlem resident, respectively, lead to an inequitable tax result. Every dollar increase in value of the QOF investment increases a QOF investor's net after tax-wealth by one dollar, while every dollar in wages earned by a typical Central Harlem resident increases that person's net after-tax wealth by about 80 cents after deducting the employee share of federal payroll taxes (7.65%) and applying the 12.19% tax rate. From a federal income tax perspective, this difference in tax treatment disproportionately increases a QOF investor's net after-tax wealth, and in turn, increases the existing racial wealth gap. Thus, the QOZ program, at least with respect to Central Harlem, seems unfair from a critical tax equity perspective, because it is most likely to have a disproportionate negative impact on Black and Hispanic QOZ residents.

There are currently 8,764[41] QOZs covering 12 percent of census tracts in the United States.[42] If QOZs across the county are like Central Harlem, the QOZ program could have a detrimental impact on racial equity nationally. It is undeniable, however, that QOF investors risk

[40] *See* Davis et al., *supra* note 4, at 6.
[41] *See Opportunity Zones*, *supra* note 1.
[42] OFF. OF THE COMPTROLLER OF THE CURRENCY, U.S. DEP'T OF THE TREASURY, *Opportunity Zones,* COMMUNITY DEVELOPMENTS FACT SHEET 1 (Aug. 2020), https://www.occ.gov/publications-and-resources/publications/community-affairs/community-developments-fact-sheets/cd-fact-sheet-opportunity-zones.pdf.

their wealth by investing in QOZs, which are, by definition, economically distressed areas. It would, thus, follow that if no tax incentive is provided to investors, they would be less likely to invest in QOZs. But while QOF investors should be compensated for their risk as an incentive to invest in QOZs, the benefits of the QOZ program should not be so one-sided in favor of QOF investors.

Proponents of the QOZ program argue that the program has caused significant investment in QOZs, spurring economic activity, and creating jobs in those communities.[43] In other words, proponents of the program argue that while the tax benefits of the program are allocated to QOF investors, substantial non-tax benefits are allocated to members of the QOZ communities mainly because of the capital invested by the QOF investors. However, the QOZ tax rules do not have any mechanism to measure whether any benefits of the program are allocated to the members of the communities in which the QOZs are located.[44] Given that the first 10-year holding period is set to expire on January 1, 2028, such that the first year in which QOF investors will be subject to the 0% effective tax rate is 2028, the Treasury Department should prioritize the collection and synthesis of tax-based demographic data with respect to QOZs. These data should measure the benefits (both tax and non-tax) of the QOZ program that are allocated among QOF investors and QOZ residents to determine whether the program advances racial equity among taxpayers.

The capital gain requirement of the QOZ program highlights one instance where traditional tax policy analysis alone is not sufficient to show how the tax system can disproportionately impact

[43] *See* Andrew Gordon, *Effectiveness of the Opportunity Zone Program*, CORNELL REAL ESTATE BLOG (Feb. 21, 2020), https://blog.realestate.cornell.edu/2020/02/21/effectiveness-of-the-opportunity-zone-program/.

[44] *See* Lisa Christensen Gee & Lorena Roque, *Opportunity Zones Bolster Investors' Bottom Lines Rather than Economic or Racial Equity*, INST. ON TAX'N & ECON. POL'Y 3 (Dec. 2019), https://itep.sfo2.digitaloceanspaces.com/121219-Opportunity-Zones-Bolster-Investors%E2%80%99-Bottom-Lines-Rather-than-Economic-or-Racial-Equity.pdf.

groups of taxpayers. To be sure, traditional tax policy should not be discarded, but can be amplified by critical tax theory. The schism between traditional tax policy and critical tax theory exists not because the two schools of thought are inherently inconsistent, but because of the dearth of reliable tax-based demographic data, the lack of which can sometimes be misunderstood as absence of evidence of racial inequity. By no means is the above analysis comprehensive, but it highlights the need for tax-based demographic information that is necessary for performing critical tax equity analysis. By adding an additional consideration – taxpayer demographics – to traditional tax policy, critical tax theory provides a lens that could help us see what traditional tax policy does not. Perhaps with more comprehensive tax-based demographic data, a critical tax lens might show that the QOZ program is advancing racial equity, including in Central Harlem. But without such data, it will be difficult, if not impossible, to measure the true impact of the QOZ program on racial equity.

WHAT DOES A SHADOW WEIGH- NOTHING OR EVERYTHING? AN ANALYSIS OF SHADOW TRADING

By: Ashini Dias

INTRODUCTION

Insider trading has been an issue for corporate America since as early as 1909 when the Supreme Court first tried to formulate and define the bounds of this new activity.[1] The U.S. Securities and Exchange Commission ("SEC") was able to pursue insider trading in the United States in 1934.[2] Over the last fifty years and more, the United States has seen notable cases alleging insider trading. These cases include Enron, Galleon Group, and even Martha Stewart's activity with ImClone.[3] The idea of insider trading has pervaded our society and become an infamous practice. As the courts and legislators have learned more about the practices and intricacies of insider trading, the laws regarding this illegal activity have also grown and expanded with time. Changes in leadership at the Securities and Exchange Commission may also be a contributing factor for deciding the metes and bounds of insider trading and how charges can be brought. In 2021, the SEC pursued a new theory of "shadow trading".[4] This theory may have blown the doors wide open on what we once knew about the illegal practice.

In August 2021, the Securities Exchange Commission filed a case with charges against Matthew Panuwat for violating section 10(b) of the Securities Exchange Act of 1934, specifically

[1] *Timeline: A History of Insider Trading*, THE NEW YORK TIMES (Dec. 6, 2016), https://www.nytimes.com/interactive/2016/12/06/business/dealbook/insider-trading-timeline.html?mtrref=undefined&assetType=PAYWALL
[2] *Federal Securities Law: Insider Trading,* CONGRESSIONAL RESEARCH SERVICE (Mar. 1, 2016), https://sgp.fas.org/crs/misc/RS21127.pdf
[3] *Notorious insider trading cases,* CNBC (Jun. 2, 2014), https://www.cnbc.com/2010/11/23/famous-insider-trading-cases.html
[4] *SEC Pursues "Shadow Trading" Insider Trading Case,* JDSUPRA (Aug. 25, 2021), https://www.jdsupra.com/legalnews/sec-pursues-shadow-trading-insider-6509820/

rule 10b-5 of Section 10(b) of the Securities Exchange Act of 1934 "gave the SEC power to enact rule against manipulative and deceptive practices in securities trading".[5] Section 10(b) became extremely popular to use during the 1960s.[6] Congress also "gave the SEC the authority and flexibility to create and revise rules, in the hope that it could effectively deter and punish manipulative and deceptive practices."[7] In essence, the claim in this case is that Matthew Panuwat is liable for participating in a new form of misconduct called shadow trading. Many corporations and law firms are now scrambling to make sense of this new theory of shadow trading and how best to inform their clients on employer contracts moving forward.[8]

The question now remains- will shadow trading be a new approach to insider trading that the courts will accept? Will it be considered government overreach for the SEC to charge defendants with a much more expansive scope of insider trading, or is this just the next logical step for the SEC to pursue? This note will try to answer these questions in the affirmative- it is likely that this new theory of insider trading will withstand scrutiny in court and be a new expansion for the illegal practice.

Part I will first analyze the general background of insider trading. This will include an explanation of the four main theories that are used to explain and understand this illegal practice. Part I will also take a closer look at the Securities Exchange Act of 1934 and specifically Rule 10(b)-5.

---

[5] *Securities Exchange Act Rule 10b,* FINDLAW (Jun. 20, 2016), https://www.findlaw.com/consumer/securities-law/securities-and-exchange-act-rule-10b.html

[6] Troy Cichos, Note, *The Misappropriation Theory of Insider Trading: Its Past, Present, and Future,* 18 SEATTLE L. REV. 389, 391 (1995).

[7] *Id.*

[8] Craig Warkol, *SEC Charges Novel Insider Trading Cases and Shines a Spotlight on 'Shadow Trading',* SCHULTE ROTH & ZABEL (Aug. 19, 2021), https://www.srz.com/images/content/1/8/v2/180346/081921-SRZ-Alert-Spotlight-on-Shadow-Trading.pdf.

Part II will delve into the specifics of *SEC v. Panuwat*. This will include an overview of the theory under which the case has been brought.

Part III will analyze cases that have used the same theory to determine whether *Panuwat* will be able to withstand court scrutiny.

Part IV will discuss the possible future of insider trading and what could happen if *Panuwat* is deemed acceptable in court.

Finally, Part V will outline how companies and firms can better advise their clients on what to include in employment contracts in an effort to avoid potential shadow trading liability. This Part will compromised an analysis of current client memos and other resources.

PART I: The Beginning of Insider Trading

A. Overview of the Securities Exchange Act of 1934

Since most of the theories of insider trading revolve around the Securities Exchange Act of 1934, it will be helpful to first give an overview of the Act and more specifically the Section 10(b)-5 provision. The Securities Exchange Act was created in 1934 to promote transparency in the sale of securities.[9] The main goal of this legislation was to regulate the market after the catastrophic stock market crash of 1929, which President Franklin Roosevelt believed was caused by shady and unregulated trading at the time.[10] The Act gives the SEC "the power to register, regulate, and oversee brokerage firms, transfer agents, and clearing agencies as well as the nation's securities self regulatory organizations".[11] This gives the SEC authority over organizations such as the New York Stock Exchange, NASDAQ market, and even the Financial Industry Regulatory

[9] Jeffrey Miron, *An Economic Defense of Insider Trading,* THE ATLAS SOCIETY (Feb. 12, 2012), https://www.atlassociety.org/post/an-economic-defense-of-insider-trading
[10] *Id.*
[11] *The Laws That Govern the Securities Industry,* INVESTOR.GOV, https://www.investor.gov/introduction-investing/investing-basics/role-sec/laws-govern-securities-industry#secexact1934

Authority, which is also known as FINRA.[12] Even with this broad and expansive authority, the SEC has regularly tried to push the boundaries of their authority to pursue a diverse portfolio of insider trading charges against a larger group of potential defendants.[13]

Though Section 10b-5 of the Exchange Act is generally an antifraud section, it is also used illuminate what the specific practice of insider trading entails.[14] This provision expressly prohibits anyone from engaging "in any act, practice, or course of business which operates or would operate as a fraud or deceit upon any person, in connection with the purchase or sale of any security".[15] This is the go-to provision for the SEC to bring claims against defendants for anything related to insider trading.[16] Additionally, Section 10(b)-5 allows not only the SEC to bring charges against companies, but also private citizens to file lawsuits against companies and other individuals that might be in violation of this provision.[17] The Supreme Court has even gone so far to hold that Section 10(b)-5 gives the SEC authority to bring insider trading charges against third-party actors; these actors were previously considered outside of the scope of the provision.[18] Many corporations, organizations and firms actively provide information to clients and other officers on how best to avoid liability in possible 10(b)-5 claims.[19]

B. Definition of Insider Trading

[12] *Id.*
[13] Jonathan Macey, *Insider Trading and the Supreme* Court, HOOVER INSTITUTION (Sep. 26, 2016), https://www.hoover.org/research/insider-trading-and-supreme-court
[14] 17 CFR 240.10b-5
[15] *Id.*
[16] *Rule 10b5,* AMERICAN BAR ASSOCIATION (Jan. 20, 2021), https://www.americanbar.org/groups/business_law/publications/the_business_lawyer/find_by_subject/buslaw_tbl_mci_rule10b5/ (explains the 388 cases against foreign issuers).
[17] *SEC Rule 10b-5,* GIBBSLAWGROUP, https://www.classlawgroup.com/securities-fraud/laws/sec/rule-10b-5/
[18] Tobi Richards, *Supreme Court provides new lens for evaluating Rule 10b-5 liability,* THOMSON REUTERS (May 7, 2019), https://tax.thomsonreuters.com/blog/supreme-court-provides-new-lens-for-evaluating-rule-10b-5-liability/
[19] https://content.next.westlaw.com/w-004-5011?isplcus=true&transitionType=Default&contextData=(sc.Default)&firstPage=true

Insider trading can be defined as "buying or selling security, in breach of a fiduciary duty or other relationship of trust and confidence, on the basis of material, nonpublic information about the security".[20] This illegal practice can come in a multitude of shapes and forms. The definition for insider trading remains broad which results in many questions on what specific situations can constitute insider trading. Common cases are those that are brought against officers or directors of companies for trading on confidential information that have not gone public[21], banking employees who learn of and execute profitable trades from companies they do business with[22], and even political consultants who might learn of trades through government work.[23] Insider trading claims can be pursued by following one of three popular theories, and one that is just emerging. These are the classical theory, tipper-tippee theory, misappropriation theory, and the outsider trading theory.[24]

1. Definition of Classical Theory. The classical theory of insider trading has been qualified as the basis of this practice by the courts for over three decades.[25] This view defines insider trading in instances where a corporate insider trades stocks in her own company which she learned through nonpublic and confidential information that was meant to be private.[26] The definition of "corporate insiders" can range from officers, directors, to general employees of the company.[27] "This theory

[20] *Supra* note 12.
[21] *Id.*
[22] *Id.*
[23] *Id.*
[24] Bradley Bondi & Steven Lofchie, *The Law of Insider Trading: Legal Theories, Common Defenses, And Best Practices For Insuring Compliance,* CADWALADER (2011), https://www.cadwalader.com/uploads/books/e5aa079bcd4be2816c4c330aaa24f380.pdf
[25] Zachary Gubler, *A Unified Theory of Insider Trading Law,* 105 GEO. L.J. 1225 (Jun. 2017), https://plus.lexis.com/document/?pdmfid=1530671&crid=16e23ec3-1538-4d31-b6ce-406452e7bf37&pddocfullpath=%2Fshared%2Fdocument%2Fanalytical-materials%2Furn%3AcontentItem%3A5NV1-DY10-02BM-Y3FR-00000-00&pdcontentcomponentid=7337&pdteaserkey=&pdislpamode=false&pdworkfolderlocatorid=NOT_SAVED_IN_WORKFOLDER&ecomp=-t4hk&earg=sr0&prid=6b884698-a5ba-4c0a-b861-cd81472747aa
[26] *Id.*
[27] *Classical Theory: Breach by a Corporate.* Insider, WILLKIE COMPLIANCE, https://complianceconcourse.willkie.com/resources/insider-trading-us-the-classical-theory

"covers situations in which a company executive, board member, or agent, such as an investment banker, trades in the company's securities or in the securities of a potential deal partner prior to the release of news concerning a significant event".[28] These significant events can include news of a tender offer, merger, or earnings announcement.[29]

In the classical theory, a rule 10(b)-5 violation exists only when this "corporate insider purchases or sells securities on the basis of material, non-public information".[30] Criminal liability can be extended to insiders who owed some level of fiduciary duty to their shareholders or board of directors. Insider trading charges can be brought civilly or administratively by the SEC, and criminally by the DOJ. Under this theory, fiduciary duty is breached when a corporate insider uses the private information to create a personal and advantageous trade.[31] The corporate insider is wrong because of the failure to disclose information about the profitable trade.[32]

The fiduciary duty requirement means that all employees of a corporation should put the interests of the corporation first before their own, especially in terms of company finances.[33] In *Guth v. Loft*, the court held that "corporate officers and directors are not permitted to use their position of trust and confidence to further their private interest".[34]

A seminal case and one that is the most popular that helped to clairfy the classical theory was *Chiarella v. U.S.*[35] In *Chiarella*, a financial printing press hired a "markup man" to help in one of their offices.[36] Chiarella had to handle a variety of office documents, five of which were

[28] *Supra* note 24.
[29] *Id.*
[30] Nic Heuer, Articles and Notes Editor, Les Reese, Annual Survey Editor, and Winston Sale, Articles and Notes Editor, of the *American Criminal Law Review*, in the Spring 2007 article "Securities Fraud", *What is Insider Trading?,* BRITANNICA (Oct. 17, 2008), https://insidertrading.procon.org/view.answers.php?questionID=001026
[31] *Id.*
[32] *Supra* note 25.
[33] *Fiduciary Duty,* CORNELL LAW SCHOOL, https://www.law.cornell.edu/wex/fiduciary_duty
[34] Guth v. Loft, 5 A.2d 503, 510 (Del. 1939).
[35] Chiarella v. U.S., 445 U.S. 222 (1980).
[36] *Id*. at 224.

related to announcements of bids for company takeovers that the organization was interested in pursuing.[37] When Chiarella received these documents, the corporation names and other confidential information were hidden with blank spaces and also included false names.[38] However, Chiarella was able to deduce the names of the corporation and purchased stock in these companies.[39] Chiarella was able to trade stocks of the target companies ahead of the public announcements, which earned him over $30,000 in just 14 months.[40] The Court disagreed with the Court of Appeals and found that Chiarella was not really an employee of the printing press, and therefore had no fiduciary duty to even breach.[41] Even though the Court eventually reversed Chiarella's insider trading charges, this case was important because it highlighted that the SEC can only successfully charge defendant's if there is a breach of some sort of fiduciary duty to shareholders or a board of directors.[42] This bolstered the classical theory point of view and affirmed the idea that insider trading under this theory can only be a crime against actual employees, and not "middle men" like what Chiarella was considered.[43]

Though the classical theory proved to be a source for many successful insider trading cases, following this method still left many gaps in liability that were allowing many insider trading situations to fly under the radar of justice.[44] One of the failures is that insider trading under this theory does not apply to "impersonal markets", so in this case someone who trades on the stock of

[37] *Id.*
[38] *Id.*
[39] *Id.*
[40] *Id.*
[41] *Id* at 232.
[42] *Id*.
[43] *Id.*
[44] *Supra* note 25.

a public company cannot be held criminally liable.[45] Additionally, passing information on to a third party to complete a trade would mean that the third party cannot be liable.

2. Definition of Tipper-Tippee Theory. The next theory that aims to make sense of insider trading is known as the tipper-tippee theory. Under this framework, there are at least two people involved in an illegal trade.[46] The person who knows of the confidential information is known as the "tipper" and the person who gets that confidential information is known as the "tippee".[47] There are four elements that a claim under this theory must meet to successfully impose liability: 1) the tipper has "breached his fiduciary duty to the shareholders by disclosing the [material nonpublic] information to the tippee, 2) the tippee "knows or should know that there has been a breach", 3) the tippee uses the information in connection with a securities transaction, and 4) the tipper receives some material benefit in return.[48] The fourth element can be read broadly, and a gain can be defined as anything from a minor profit to a "potential" to earn greater in the future.[49] The tipper-tippee angle also makes it illegal for a tipper to give information to a friend or relative.[50] This theory clearly captures more people than the classical theory did and was a more expansive view that the SEC could pursue.

Dirks v. SEC was a leading case that helped to solidify the tipper-tippee theory.[51] In this case, Dirks received information about potential fraud at Equity Funding and went to do an investigation.[52] While conducting the investigation, he was telling a number of clients and other

[45] *Id.*
[46] Sari Rosenfeld, *The Ever-Changing Scope of Insider Trading Liability for Tippees in the Second Circuit*, MICH. BUS. AND ENT. L. REV. 401 (2019).
[47] *Id.*
[48] *Id.*
[49] *Id.*
[50] *Id.*
[51] Dirks v. SEC, 436 U.S. 646 (1983).
[52] *Id.*

financial officers about his findings.[53] The stock price for Equity Funding fell from $26 to $15 in the time period that Dirks was doing his investigation, and the SEC began to investigate his involvement with the company once the Wall Street Journal had .published an article explaining the corporate fraud.[54] The SEC eventually charged Dirks with violation of Section 10(b) of the Exchange Act and claimed that he was guilty of insider trading for sharing confidential information during the investigation.

The Supreme Court eventually reverses the charges against Dirks and hold that there was no violation of duty that is owed directly to shareholders.[55] The court affirms the elements of insider trading on a tipper-tippee theory by highlighting that "the tippers received no monetary or personal benefit for revealing Equity Funding's secrets, nor was there purpose to make a gift of valuable information to Dirks", which is an essential element to a claim of insider trading.[56]

Another popular case that used the tipper-tippee theory was *U.S. v. Martoma*.[57] Here, Martoma was involved with an insider trading scheme dealing with two pharmaceutical companies that were experimenting with a new drug.[58] Martoma acquired insider information about the success of the company's stocks that was relayed to him by someone working closely with the prescription drug trials.[59] Martoma tried to get a reversal of his charge by stating that the "personal benefit" element needed proof that there was also a "meaningfully close personal relationship"

[53] *Id.*
[54] *Id.* at 652.
[55] *Id.* at 665.
[56] *Id.* at 666.
[57] *Second Circuit Again Holds That Tipper/Tippee Liability Can Arise from a Gift of Inside Information Even Without a Close Personal Relationship,* PROSKAUER (Jun. 28, 2018), https://www.proskauer.com/alert/second-circuit-again-holds-that-tipper-tippee-liability-can-arise-from-a-gift-of-inside-information-even-without-a-close-personal-relationship, United States v. Martoma, 894 F.3d 64 (2018).

[58] Martoma, 894 F.3d at 69.
[59] *Id.* at 70.

between the tipper and tippee.[60] Martoma's argument is that since this definition was left out of the jury instruction, the charges should therefore be dismissed.[61]

The Second Circuit finally affirmed the ruling by reanalyzing the *Dirks* standard for what constitutes a "close relationship".[62] The court explained that "personal benefit" was read broadly and many situations could satisfy this element.[63] Personal benefit could range from giving information to "friends from college" or exchanging an iPhone for entry to an investment club where a tippee could have the opportunity to learn about profitable trades.[64] The court was satisfied "that the personal benefit element can be met by evidence that the tipper's disclosure of inside information was intended to benefit the tippee".[65]

3. Definition of Misappropriation Theory. The theory that is given most importance in this note is the misappropriation theory of insider trading. This theory aims to take a very board approach to the definition and understanding of the 10(b)-5 provision of the Exchange Act.[66] This theory "has evolved into the SEC's preeminent prosecutorial weapon for combating fraudulent trading practices".[67] Under misappropriation, a defendant can be found guilty of committing some act of insider trading if she violates any kind of duty to any person with information that was meant to be private and confidential.[68] This is clearly a very broad reading of the 10(b)-5 provision because this theory now catches many situations that previously could go under the radar. The misappropriation theory is different from the classical theory because the definition of insider

[60] *Id.* at 68.
[61] *Id.*
[62] *Id.* at 78.
[63] *Id.* at 74.
[64] *Id.*
[65] *Id.* at 75.
[66] *Supra* note 7.
[67] Joseph Humke, *The Misappropriation Theory of Insider Trading: Outisde the Lines of Section 10(b)*, MARQUETTE L. REV. VOL. 80, (1997) https://scholarship.law.marquette.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1508&context=mulr
[68] *Supra* note 67.

trading is no longer confined only to those who breach a duty owed to shareholders of a company.[69] Liability can attach to an individual in a wide range of scenarios that are not just between an insider of a corporation.[70]

In short, any instance of trading that results in an individual gain with the use of material, nonpublic information in violation of some duty of trust will likely be found to be an issue of insider trading.[71] This theory became very popular in the late 1990's when the Supreme Court held that the misappropriation theory was a valid way to impose 10(b) liability on a defendant to impose criminal charges in the prominent case *United States v. Hogan*.[72] This case opened the gates for the SEC to use the misappropriation theory in a variety of ways that were previously not possible before the emergence of this new avenue for insider trading charges.

This is the theory that the current shadow trading case is relying on in *SEC v. Panuwat* in an effort to bring criminal charges against the defendant using powers of the SEC authority.[73] The rest of the note will go into detail on how the SEC is trying to apply misappropriation and other noteworthy cases that have successfully used misappropriation in court which can be used as precedent when analyzing the current shadow trading case. The cases that will be able to successfully apply the misappropriation theory might change once a court makes a decision on the shadow trading case.

---

[69] *Id.*
[70] *Id.*
[71] *Id.*
[72] Keith Simon, *The Misappropriation Theory: A Valid Application of 10(b) to Protect Property Rights in Information,* J. OF LAW AND CRIMINOLOGY VOL. 8, (!998), https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=6968&context=jclc
[73] *Supra* note 5.

4. Definition of "Outsider Trading" Theory. It is also worth mentioning a less prominent theory of insider trading which is known as the "outsider trading" theory.[74] This theory can also be called the "affirmative misrepresentation" theory.[75] This view can be considered the most expansive approach since it allowed for an insider trading charge to proceed without the presence of a fiduciary duty.[76] This theory first came to light in 2009 with the case *SEC v. Dorozhko*.[77] In this case, Dorozhko was a computer hacker who downloaded an online trading account around the same time that the company IMS Health announced it would release certain financial data to the public.[78] Dorozhko was able to hack into the servers and learn about the stocks minutes before the public announcement, and completed a trade that earned him almost $250,000 overnight.[79] The court explains that deceptive practices require a breach of fiduciary duty to hold someone liable. However, the court argues that computer hacking is a "distinct species of fraud" that is illegal regardless of whether or not a fiduciary duty can be shown.[80] The case was remanded to determine whether or not Dorozhko's actions could be considered deceitful, but this highlighted a new avenue where defendants could be liable even without the existence of some kind of fiduciary duty.[81] This theory has not been tested much and is mostly reserved for very specific instances that include computer hacking. One author explains that "one can envision the affirmative misrepresentation concept extended to investment interactions where institutional investors are accused of tricking market participants into sharing material nonpublic information".[82] Nevertheless, it is still an

[74] *Supra* note 25.
[75] *Id.*
[76] *Id.*
[77] SEC v. Dorozhko, 574 F.3d 42 (2009).
[78] *Id.* at 44.
[79] *Id.*
[80] *Supra* note 66.
[81] *Id.*
[82] *Id.*

interesting approach that describes more modern day issues of computer hacking and shows once again an expansive view of insider trading that is trying to catch more activity under this definition.

Part II: The SEC's new attempt at "shadow trading": *SEC v. Panuwat*

In the case at issue, the SEC is looking to charge a private individual with what experts in the field of insider trading have come to call "shadow trading".[83] This can most certainly be seen as an expansion of what constitutes insider trading and an effort on behalf of the SEC to charge more private individuals with committing this particular activity.[84] The case goes on to describe the defendant in detail, his position at the company he worked for, the alleged crime, and the course of action that the SEC is urging the district court to take.[85] The SEC breaks up its complaint into various sections that will be detailed here.

On August 17, 2021 the SEC filed this complaint against Matthew Panuwat in the United States District Court in the Northern District of California.[86] The SEC explains that it is bringing this action and that it is permitted to under Section 21A of the Exchange Act.[87] Section 21A explains that if the Commission believes someone to have purchased or sold stock with material and non-public information or in violation of a duty of good faith, then the organization will be allowed to seek redress in the district court.[88] This section also explains that the district court will then have jurisdiction to impose a penalty on the party that has committed the crime.[89] The SEC is additionally making clear in the complaint that interstate commerce was being affected by

[83] https://www.wlf.org/2021/09/07/wlf-legal-pulse/sec-takes-a-crack-at-expanding-misappropriation-theory-to-shadow-insider-trading/
[84] *Id.*
[85] https://www.sec.gov/litigation/complaints/2021/comp-pr2021-155.pdf
[86] *Id,* cover page.
[87] *Id* at page 3, section 8
[88] 15 U.S.C. §78u-1
[89] *Id.*

defendant's actions.[90] The complaint alleges that the defendant "directly or indirectly, made use of the means and instrumentalities of interstate commerce or of the mails in connection with the acts, transactions, practices, and courses of business alleged in this complaint".[91]

The complaint goes into detail on the defendant's background and history up until the point of being charged with a new expansion of insider trading. Panuwat was a resident of California where he was working for a company called Medivation in the business development sector of the corporation.[92] Panuwat was considered to be an "expert" in the biopharmaceutical industry.[93] The SEC explains this is because Panuwat held a biology related undergraduate and graduate degrees, as well as an MBA that would make him a professional in a business that is comprised of sales and productions of pharmaceuticals.[94] The complaint stresses that Panuwat is a professional and knew exactly what he was doing when he committed the insider trading crime- he spent over 15 years "in the biopharmaceutical industry, including eight years in the global healthcare investment banking division of a top investment bank and employment in business and strategic development at several biopharmaceutical companies".[95] In addition to his work with Medivation, he was also an executive at another publicly-traded company that also dealt with biopharmaceuticals.[96]

The complaint briefly explains that Medivation was a biopharmaceutical that focused their business on oncology efforts.[97] Medivation was acquired by Pfizer in 2016 and was then considered a more global and international reaching company.[98] Panuwat traded stocks in Incyte,

[90] https://www.sec.gov/litigation/complaints/2021/comp-pr2021-155.pdf page 3, section 10
[91] *Id*.
[92] *Id* page 3, Section 13.
[93] *Id* page 3, section 14.
[94] *Id.*
[95] *Id.*
[96] *Id.*
[97] *Id* page 3 section 15.
[98] *Id.*

which was another oncology specialized biopharmaceutical company which had a principal place of business in Delaware.[99] This company had also registered with the SEC in accordance with required business practices and was also listed on the Nasdaq.[100]

Panuwat's job function at Medivation included various responsibilities, some of them being to create business development reports for the company's Chief Financial Officer.[101] This extensive job responsibility also included closely tracking "the stock prices, drug products, and development pipelines of other biopharmaceutical companies, including Incyte, as well as merger and acquisition activity in the biopharmaceutical industry".[102] Because Panuwat dealt with such sensitive and confidential information on a consistent basis, part of his employment agreement included agreeing to not make use of any of this information.[103] The only exception to this rule was that he was permitted to use information that would be beneficial to Medivation as a whole.[104] Additionally, Panuwat signed an insider trading policy which "prohibited employees from personally profiting from material nonpublic information concerning Medivation by trading in Medivation securities or the securities of another publicly traded company."[105] This might have been Medivation's way of making sure employees did not face insider trading liability. It seems that the policy was left to be intentionally general to catch any activity that was not expressly stated anywhere else in the agreement. In 2016, there had been discussion that Medivation might be acquired by a specific French company.[106] In regards to this effort, Panuwat was part of the

[99] *Id* page 4 section 16.
[100] *Id.*
[101] *Id* page 4 section 17.
[102] *Id.* page 4 section 18.
[103] *Id.* page 4 section 19.
[104] Id.
[105] *Id.* page 5 section 20.
[106] *Id.* page 5 section 21

discussion with various investment banks about the options for acquisiton which included a discussion on other comparable businesses that could be related to Medivation in a possible merger to make the acquisition option seem more successful.[107] Much of this discussion and extensive meeting was focused around Incyte being one of the only comparable corporations to Medivation and Panuwat knew that there were very few other companies that could be comparable.[108] Panuwat had further knowledge that when a similar deal happened in the past with a mid-size company that focused on oncology, both Medivation and Incyte (which were considered comparable corporations) saw quite a significant increase in their stock prices.[109] Throughout the entire discussion and analysis of these comparable companies, Panuwat was around sensitive and confidential information that included potential share prices for the impending acquisition. This involvement included "coordinating Medivation's responses to various due diligence requests during that time and participating in meetings of Medivation's board of directors concerning Medivation's strategic alternatives with respect to a potential merger".[110]

In August of 2016, Panuwat was party to a meeting of the Medivation Board of Directors where the discussion centered around final bids and stock prices for an imminent acquisition of the company.[111] Panuwat was in possession of certain letters that included this sensitive information prior to the meeting as well.[112] After the finalization of the acquisition and before it was available to the public, Medivation sent out an email that included Panuwat on the premium to the share price.[113]Almost immediately, Panuwat used his "personal brokerage account from his

[107] *Id.* page 5 section 22.
[108] *Id.*
[109] *Id.*
[110] *Id* page 6 at 25.
[111] *Id* page 7 at 28.
[112] *Id.*
[113] *Id.* page 8 at 33.

work computer and purchased 578 Incyte call option contracts with strike prices of $80, $82.50, and $85 per share…".[114] Panuwat was not relying on any earnings report or other information from Incyte, but instead was just relying on the anticipation that Incyte's stock price would increase once the public disclosure of Medivation's acquisition had been announced.[115] Panuwat did not tell anyone at the company that he bought these stocks.[116] He similarly did not receive permission or clearance from anyone at Medivation to complete this purchase of stock which the SEC considered to be self-serving and "in breach of his duty of trust and confidence…".[117] A few days after the purchase of these options, Medivation publicly announced the acquisition of the company by Pfizer with a 21.4% premium over its closing price per share on Friday.[118] It is also interesting to note that in the report that was conducted by investment bankers, Incyte was named and included as a similarly publicly-traded company in regards to fairness opinions on what the price of the new Medivation stock would be.[119]When the market opened, Medivation's stock price "climbed materially to $80.62, a 20% increase over the prior trading day's closing price of $67.16".[120] As expected, Incyte's stock price also materially increased and eventually reached a high of $84.39.[121] In the end, Panuwat saw a profit of $107,066 through the purchase of Incyte stock prior to the public announcement of Medivation's acquisition.

The SEC is trying to allege that Panuwat's actions and purchase of Incyte's stock options was in violation of Section 10b of the Exchange Act.[122] As previously discussed, Section 10b and

[114] *Id.*
[115] *Id.*
[116] *Id.* page 8 at 34.
[117] *Id.*
[118] *Id.* page 9 at 36.
[119] *Id* page 8 at 35.
[120] *Id.* page 9 at 36.
[121] *Id.* page 9 at 37
[122] *Id.* page 9 at 39.

more specifically Rule 10b-5 is part of a general provision that prohibits fraud in connection to the purchase or sale of securities.[123] Illegal conduct can include making untrue statements or omitting certain facts in order to mislead or make a stock option seem more desirable.[124] Additionally, the SEC argues that if he is not stopped, that Panuwat will continue to violate Section 10b and especially as a director of another publicly-traded biopharmaceutical company he is constantly being exposed to sensitive information that will make him commit fraud once again.[125] The SEC is asking the court to order Panuwat to "pay a civil monetary penalty…" [Not sure why you need a quote here] in addition to prohibiting him from ever serving as an officer or directory of a corporation that sells securities on the market.[126]. In this situation, Incyte would be considered the "shadow" company whose actions are likely to mimic the corporation that defendant is an actual insider to.[127] The SEC is taking the misappropriation theory approach to this action against Panuwat which is already quite a broad understanding of what constitutes insider trading.[128] The decision in this case will determine how far the misappropriation theory can extend to cases that are not typically considered to be insider trading, along with including many more defendants in possible suits for attempting to defraud in connection with the purchase or sale of securities.

Part III: Misappropriation Precedent

A. *United States v. O'Hagan*

The seminal case brought under the misappropriation theory of insider trading was that of *United States v. O'Hagan.*[129] Following the decision and success in this case, the Supreme Court cleared

[123] *Id.* page 9 at 40.
[124] *Id.*
[125] *Id.* page 10 at 41.
[126] Id., prayer for relief section.
[127] *Supra* note 1.
[128] *Id.*
[129] United States v. O'Hagan, 521 U.S. 642 (1997).

the path for more insider trading cases to be pursued by the SEC under this more expansive theory. Particularly in this case, the court was looking to address two questions, but one is the most important at issue which is as follows: "is a person who trades in securities for personal profit, using confidential information misappropriated in breach of a fiduciary duty to the source of the information, guilty of violating §10(b) and Rule 10b-5?"[130] This case was decided in 1997 following a long history of the SEC trying to expand the limits of insider trading.

In this case, the Defendant James O'Hagan was a partner at a major law firm in Minnesota.[131] O'Hagan had been a long term employee of the firm and had been involved in a multitude of other similar business transactions. In 1988, a company by the name of Grand Metropolitan PC hired O'Hagan's law firm to represent the company "to represent Grand Met regarding a potential tender offer for the common stock of the Pillsbury Company…"[132] O'Hagan was not directly tied to the Grand Met business deal and did not do any work regarding the tender offer relating to Pillsbury Company.[133] In September, two months after Grand Met had retained the law firm as counsel, the firm withdrew representation.[134] Shortly after in October, Grand Met announced the stock buying option that they had decided on for Pillsbury Company which became available to the public.[135] The issue in this case is that before O'Hagan's firm had stopped representing Grand Met, O'Hagan decided to purchase call options for Pillsbury stock.[136] According to Justice Ginsburg's opinion for the case, "each option gave him the right to purchase 100 shares of Pillsbury stock … Later in August and September, O'Hagan made additional

[130] *Id* at 647.
[131] *Id.*
[132] *Id.*
[133] *Id.*
[134] *Id.*
[135] *Id.*
[136] *Id.*

purchases of Pillsbury call options."[137] In the month of September, O'Hagan had once again purchased shares of Pillsbury stock, where the additional purchase was made at around $39 a share.[138] Following the public announcement of the Grand Met tender offer, the price of the stock rose exponentially to $60 a share.[139] O'Hagan was then able to sell his call options and stock and finally had a cash out which allowed him to make a profit of more than $4.3 million dollars.[140]

The profit price was exponential, which alerted the SEC to conduct an investigation into potential insider trading.[141] Following a thorough investigation into the actions surrounding the Grand Met and the stock options for Pillsbury company, the SEC charged O'Hagan with a massive 57-count indictment.[142] The crux of this indictment was that the defendant "defrauded his law firm and its client, Grand Met, by using for his own trading purposes material, nonpublic information regarding Grand Met's planned tender offer".[143] The SEC pursued this case against O'Hagan by additionally claiming that he used the money he earned from his insider trading deal to conceal other instances of embezzlement that he had been involved in and other issues dealing with client trust funds.[144] In the end, a jury convicted on all 57 counts that he was charged with and O'Hagan was sentenced to 41 months in prison.[145]

The case went before the Supreme Court on appeal, where the Court reversed the Eighth Circuit reversal of the conviction.[146] The Eighth Circuit reversed all of the 57 convictions

[137] *Id.*
[138] *Id.* at 648.
[139] *Id.*
[140] *Id.*
[141] *Id.*
[142] *Id.*
[143] *Id.*
[144] *Id.*
[145] *Id.* at 649.
[146] *Id.*

following a divided panel.[147] The reason for the reversal was that the Circuit court believed that the insider trading charges could not be predicated on the misappropriation theory.[148] The major importance of this case is that the Supreme Court reversed this holding and said that the misappropriation theory is a legal and valid way for the SEC to pursue insider trading.[149] The Court explained that §10(b) of the Exchange Act provides for this new and burgeoning theory of misappropriation to hold, which would eventually expands the SEC's power to catch more defendants under insider trading.[150]

The Court gives a brief summary of the misappropriation theory which it validates in its holding.[151] Ginsburg's opinion explains that "a person commits fraud 'in connection with' a securities transaction… when he misappropriates confidential information for securities trading purposes, in breach of a duty owed to the source of the information".[152] When the insider is using confidential information for a personal gain, he is defrauding the principal of the "exclusive use" of that information.[153] The misappropriation theory does not necessarily require a fiduciary relationship, but premises this fiduciary duty on the "deception of those who entrusted him with access to the confidential information".[154] The court agrees with the Government's initial stance that the breach of fiduciary duty is predicated on deception or concealment of using material information to the principal, which they hold is consistent with the language in §10(b) of the Exchange Act.[155] The Court emphasized the fact that cases attempted under the misappropriation

147 *Id.*
148 *Id.*
149 *Id.* at 650.
150 *Id.*
151 *Id.* at 652.
152 *Id.*
153 *Id.*
154 *Id.*
155 *Id.* at 654.

theory must include some kind of manipulation or deception, otherwise it would fail to meet statutory requirements of the Exchange Act provision.[156] If the defendant had in contrast disclosed his decision to purchase the stock options to the principal (in this case, the firm that O'Hagan worked for) then there would be no deception and therefore no liability.[157] The opinion of the case explains "because the deception essential to the misappropriation theory involves feigning fidelity to the source of the information, if the fiduciary discloses to the source that he plans to trade on the nonpublic information, there is no 'deceptive device' and thus no §10(b) violation…".[158]

The next requirement of §10(b) states that the fraud or deception has to be "in connection with the purchase of sale of a security".[159] The court reasons this is also satisfied by the misappropriation theory.[160] The connection exists when the insider purchases or sells the security without disclosing it to the principal, not when he initially gains the material information.[161] The court explains that that the "securities transaction and the breach of duty thus coincide".[162] The insider who is misappropriating the material and non-public information gains the advantageous market position because of his deceitful practice or attempt to defraud a principal.[163] This action also ends up harming the public that invests in the same stocks and interferes with proper market practices.[164] In this case, the Supreme Court holds that for a defendant to be liable under the misappropriation theory, there must be evidence that shows that the defendant is using information

[156] *Id.* at 655.
[157] *Id.*
[158] *Id.*
[159] *Id*.
[160] *Id.* at 656
[161] *Id.*
[162] *Id.*
[163] *Id.*
[164] *Id.*

in order to "capitalize upon [it] to gain no-risk profits through the purchase or sale of securities."[165] If for example the insider uses the non-public information in a different way that is not related to realizing a self-serving profit, then that defendant cannot be held liable for insider trading through §10(b).[166] The court makes mention of the fact that the misappropriation theory is not meant to be a sort of catch-all tool; there are restrictions on the rule.[167] The main restriction on the rule is that it can only be used to catch "fraudulent means of capitalizing on such information through securities transactions".[168] The Court wants to ensure proper market practices and to make sure that the trading market is free from easy ways of defrauding the public and delegitimizing the purchase and sale of securities.[169]

The Court of Appeals rejected the misappropriation theory for two reasons.[170] The first reason is that the Court believed "neither misrepresentation nor nondisclosure" was require which is the way the Eight Circuit had understood the theory to mean.[171] However, the Supreme Court re-evaluates this and explains that deceptive nondisclosure is actually essential to claiming insider trading under this theory and for meeting the statutory requirements of §10(b).[172] The second reason that the Court of Appeals rejected this theory is because the court believed that there was no "in connection to the purchase or sale of any security" component of misappropriation that was required in §10(b).[173] The only way this component would be satisfied would be to show that

[165] *Id.*
[166] *Id.*
[167] *Id.*
[168] *Id.*
[169] *Id.* at 658.
[170] *Id at 660.*
[171] *Id.*
[172] Id.
[173] *Id.*

there was "only a breach to the parties to the securities transaction".[174] The Supreme Court goes the other direction on this- the argument being that 10(b) regards any security and "not to identifiable purchasers or sellers of securities".[175] The court is reading §10(b) more broadly than that of the Court of Appeals.[176]

The decision in *O'Hagan* provided some clarity around the situation of insiders trading on non-public information when there are no connections to company in question or there was no tipping involved as was the case here, where the defendant becomes known as an "outsider".[177] This expansive view of insider trading made it easier for the SEC to pursue many different types of insider trading because the court read §10(b) quite broadly than before.[178] *SEC v. Panuwat* would be a new expansion on the misappropriation theory and will be interesting to see if a court will allow a broad reading to extend to this case as well.

B. *SEC v. Cooperman*

Another notable case that proved successful was that using the misappropriation theory was *SEC v. Cooperman.*[179] Even though this was a District Court opinion, it is still important as precedent for expounding on the misappropriation theory that was laid down with *O'Hagan.* In this case, the SEC wanted to charge Leon Cooperman and his advisory firm with violating section 10(b) of the Exchange Act of 1934 and for engaging in insider trading in violation of Rule 10b-

[174] *Id.*
[175] *Id.*
[176] *Id.*
[177] Rachel Goldstein, Note, *Insider Trading and United States v. O'Hagan: The Supreme Court Reinstates Securities Fraud Convictions Based on the Misappropriation Theory, 1* CHAPMAN L. REV. 119, 123 (1998).
[178] *Id.*
[179] SEC v. Cooperman, 243 F. Supp. 3d 597 (E.D. Pa. 2017)

5.[180] Cooperman was the president, chief executive officer and majority stockholder of the advisory firm Omega, where he managed certain hedge funds and gave investment advice to a number of clients.[181] One of his clients, called Atlas Pipelines Partners (APL) received a confidential offer to purchase a certain operating facility for $720 million.[182] Cooperman began having confidential discussions with certain officers of APL regarding this purchase.[183] The information that was discussed on these calls was considered material and nonpublic, and the executives of APL believed that Cooperman "had an obligation not to use that information to trade APL securities".[184] It is important to note that there was not necessarily a contract that did not allow Cooperman's trading on APL stocks. On one phone call, Cooperman explicitly agreed not to trade on these securities.[185] However, with direction from Cooperman, the Cooperman Offshore Account and the Hedge Fund and Managed Accounts started acquiring large numbers of APL securities within a 2 week period.[186] After an agreement was finalized following the purchase of these stocks, Cooperman was notified by APL about the successful sale. Cooperman then relayed this news to an Omega consultant and told the consultant what the sale was going to be worth.[187] The consultant and Cooperman discussed the impact of this sale on Omega and eventually Cooperman directed the purchase "of additional APL securities worth approximately $620,000. Over the next week… continued to direct the purchase of APL securities".[188] Following these purchases, APL's board approved the sale and was publicly announced for the first time that it was

[180] *Id* at 600.
[181] *Id.*
[182] *Id.*
[183] *Id.*
[184] *Id.*
[185] *Id.*
[186] *Id.*
[187] *Id.*
[188] *Id.* at 601.

selling the operating facility for a final cost of $682 million.[189] As a result of this announcement, the securities that Cooperman traded on realized a value of approximately $4.09 million spread across the Offshore Account, Hedge Fund Accounts, Managed Accounts, and Family Accounts.[190] The SEC filed a complaint around three months later, relying on the misappropriation theory to find Cooperman and Omega liable for insider trading.[191] Cooperman replied to the complaint explaining that the SEC failed "to sufficiently plead an insider trading claim".[192] This reply also included an argument that the Eastern Pennsylvania District court was the wrong venue for this claim.

The court explains in the opinion that a 10(b) claim requires "heightened pleading requirements" including the "who, what, when, where and how of the events at issue".[193] Here, the SEC has met this evidentiary burden because they alleged in great detail how Cooperman's actions were deceptive and why it constitutes insider trading. It is important to note this higher standard the SEC must meet in order to even bring a 10(b) case.

The issue that the court addresses is if, under the misappropriation theory, the law requires duty of trust and confident to the executive of APL to exist when he provided the information initially to Cooperman in order to constitute insider trading. [194] The SEC argued that there is no requirement under the misappropriation theory "that an agreement not to trade precede disclosure of the confidential information, so long as a duty of trust and confidence exist **at the time the recipient trades on the information".**[195] The District court explains that legislative history deems

[189] *Id.*
[190] *Id.*
[191] *Id.*
[192] *Id.*
[193] *Id.* at 602.
[194] *Id.*
[195] *Id.*

that 10(b) was aimed at ANY manipulative or deceptive practices which would be detrimental to the interests of the investor.[196] The court goes on to say that the rule is to be read flexibly because the end goal is to maintain a fair market.[197] The rule of the misappropriation theory itself makes no mention of time, just that it was to capture illegal trading when there is deception of those who trusted him with confidential information.[198] Even if cases like this have "at least a partial source of a duty of trust and confidence suggest that the agreement may be made post disclosure".[199] Precedent shows that the Supreme Court reads 10(b) quite expansively and broadly, which is the similar view that the District Court takes to comport with previous rulings. In the end, the court sides with the SEC and rules that the misappropriation theory may include post-disclosure agreements.[200] This is consistent with the intent of Congress which the court believes was intended to target "deception that is detrimental to the rightful owner of the information, investors, and the public". [201] If the court were to rule that the misappropriation theory requires pre-disclosure agreements not to trade on securities as the only way to be held liable for insider trading, this would create an unnecessary loophole to the law that Congress did not intend.[202] If this view was accepted, corporations would be able to avoid liability which could lead to potential other illegal practices.[203] The court gives a scenario where if the trading outsider fails to get an agreement not to trade before disclosure of confidential information, this would insulate the outsider from any liability. The opinion makes clear that 10(b), whatever the language may be, is ultimately used to

[196] *Id.*
[197] *Id.*
[198] *Id.*
[199] *Id.*
[200] *Id.* at 605.
[201] *Id.*
[202] *Id.* at 606.
[203] *Id.*

avoid situations like this. As the court notes, "although defendants seek a free pass for well-timed deception, no such free pass is found in the language of Section 10(b)…"[204] Ultimately, the court decides that there is no timing requirement that an outsider could use to avoid liability and reads Section 10(b) broadly to capture a wide range of illegal activity that would constitute insider trading. The SEC's evidence of the telephone calls concerning the sale along with Cooperman's agreement not to trade on the information afterwards was enough to create a successful misappropriation claim.[205]

Part IV: Potential Implications of *SEC v. Panuwat* on Insider Trading Liability

Based on the precedent that exists before the courts today, it is likely that *Panuwat* will survive in court and Panuwat will be found guilty of committing insider trading. This means that *Panuwat* will most likely expand the misappropriation theory to include other potential shadow trading cases.

A court will be able to see that Panuwat's decision to trade on the "shadow" company's stock options would have defrauded the principal of its exclusive use of that information.[206] It is important to note that in the *O'Hagan* decision, the supreme court leaves "exclusive use of that information" quite broad. The court did not stress whether that information includes only details about the defendant's corporation (as it did in *O'Hagan*) or information that is related to the corporation, though indirectly. Considering the major discretion that courts typically award to the SEC, it is likely that a court will deem **any** information related to the purchase or sale of a corporation's securities belong exclusively to the employer corporation. Therefore, even though Panuwat was trading on a different company, because Incyte was a comparable corporation, his

[204] *Id.*
[205] *Id.* at 607.
[206] *Supra* note 154.

trading deprived Medivation of this exclusive information. *O'Hagan* describes that the misappropriation theory does not necessarily require a fiduciary duty to arise but instead explains that this duty is premised on "the deception of those who entrusted him with access to the confidential information".[207] Panuwat clearly deceived Medivation's board of directors by silently trading on Incyte's stocks. He was party to confidential information where he knew that Incyte was the only other corporation comparable to Medivation. He therefore had to have knowledge that a Medivation purchase would directly affect Incyte. Even if there was no fiduciary duty (which there clearly is since Panuwat was a direct employee of Medivation), his action would be considered deceptive. This is furthered by the fact that Panuwat signed an insider trading policy which forbade employees from trading on any security related to Medivation stocks.[208] This created some level of duty towards the employer which they relied on when looping Panuwat in for confidential board meetings regarding Medivation's future. Panuwat went against his contract as an employee which a court will likely consider enough by itself to hold Panuwat accountable for insider trading. In *O'Hagan*, the court opinion does not show that O'Hagan was required to sign any sort of insider trading policy agreement the way that Panuwat was required to. Even without this sort of contract, the court found O'Hagan guilty of insider trading. With this contract in place in Panuwat's case, it will be even harder for a court to not find an insider trading violation. If Panuwat had openly told his employer that he was looking to trade on Incyte's stocks, there would have been no deception and therefore no liability.[209]

Panuwat will also find himself in hot water under *Cooperman. Cooperman* echoed the rules following *O'Hagan*, and added that the misappropriation theory did not require that an agreement

[207] *Supra* note 155.
[208] *Supra* note 106.
[209] *Supra* note 158.

not to trade had to precede disclosure of the confidential information.[210] As long as there is a duty of confidence and trust at the time of trading on the stock, a defendant could get in trouble for insider trading.[211] Again, there was a duty instilled upon Panuwat of both confidence and trust not to trade in a way harmful to Medivation. The Board of Directors obviously brought him on because of his experience and expertise related to Medivation in order to maximize Medivation's best interest in the deal. Panuwat knew that the information was confidential and not open to the public, which proves that he further knew there was some kind of duty of confidence or trust he had to abide by. His case is made worse with the fact that there was a contract in place prior to disclosure unlike in *Cooperman*. This would further sway a court to find an insider trading violation. The *Cooperman* court stressed that no matter what the case was, the main goal of Rule 10b-5 and Section 10(b) was to promote a fair securities market.[212] Securities legislation was to be read flexibly, and provides major deference to the SEC in their actions..[213] This means that a court again is likely to find that Panuwat has committed insider trading under the misappropriation theory. It will be seen as unfair that Panuwat could trade on a very similar company, realize a major profit, and not be held liable in any way. *Cooperman* goes on to say that Congress did not mean for there to be any loopholes that would insulate a defendant from any liability by reading so closely between the lines of the legislation.[214] The court should look at the policy ramifications of allowing Panuwat to evade liability and the future possible offenses that others could engage in in the future. This might create a slippery slope that the court will want to address to deter other similar situations. Congress would have wanted Rule 10b-5 and Section 10(b) to catch any loophole that

[210] *Supra* note 196.
[211] *Id.*
[212] *Supra* note 198.
[213] *Id.*
[214] *Supra* note 203.

could affect the integrity of the securities market. This specific case is a loophole that should be closed if the court looks to follow precedent.

Adding more salt to the wound, the Northern District of California recently denied Panuwat's motion to dismiss the case.[215] The District Court is taking an expansive approach to the misappropriation theory as courts before have. The order explains that 10b-5 does not require the information about the corporation to come from the issuer itself in order to be material.[216] As long as it has a connection with any security, that could apply to a range of companies and not just the employer.[217] This District Court seems to support the SEC's broad understanding of the misappropriation and is allowing the case to move forward. The court also stressed that even though Incyte securities were not listed in the insider trading policy that Panuwat signed, this list was not supposed to be exhaustive.[218] Looking at this from a more logical perspective, allowing Panuwat to trade on Incyte securities would go against the goal of this policy that all employers were to sign.[219] Additionally, the court explains that a "reasonable investor" would deem the information important since the price of Incyte stocks increased when the deal was publicly announced.[220] This is entirely consistent with the policy considerations that the *Cooperman* court took when they explained that the end goal of the legislation was to be read flexibly.[221] Even though the court agrees that this type of case is unusual and novel, that itself is not reason to halt

[215] *District Court Denies Motion to Dismiss SEC's First 'Shadow Trading" Complaint,* KRAMER LEVIN, (Jan. 25, 2022), https://www.kramerlevin.com/en/perspectives-search/district-court-denies-motion-to-dismiss-secs-first-shadow-trading-complaint.html
[216] Cydney Posner, *SEC's "shadow trading" case survives motion to dismiss,* COOLEY PUBCO, (Jan. 20, 2022), https://cooleypubco.com/2022/01/20/shadow-trading-case-survives-dismissal/
[217] *Id.*
[218] *Id.*
[219] *Id.*
[220] *Id.*
[221] *Supra* note 198.

the SEC investigation.[222] This motion to dismiss means that the District Court is confident that there is a possibility that the SEC can expand the misappropriation theory to include other shadow trading cases. They are interested to see how this case will proceed, and support the findings that the SEC have already made in building their case. This denial of the motion to dismiss could prove significant and is a promising show that courts now would be accepting of this broad approach to insider trading.

Ultimately, it is very likely that a higher court will rule against Panuwat and in favor of the SEC. It seems that shadow trading is the next logical step in attacking insider trading cases when compared to court precedent. *Panuwat* presents a more nuanced fact pattern than what courts have seen before, but the rules from precedent seem to be intentionally broad enough to allow a legal expansion into this new field. Congress intended to promote and maintain a fair securities market and this goal will be properly served by expanding misappropriation to include shadow trading.

Part V: Advice to Firms and Corporations Moving Forward

If the court does in fact rule in favor of the SEC, that will mean that many firms and corporations will have to be more thoughtful in advising their employees and clients on insider trading rules. It will change the face of business and will allow the SEC to take a stricter and more aggressive prosecution approach. Thisi could prove quite difficult to avoid liability in a number of cases. It would be wise to implement new policies and rules in anticipation of the acceptance of shadow trading as a new approach to insider trading since it will protect employers and firms from stricter liability. This section will look at current client memos to determine what the best approach to new policy should look like. It will also try to offer other novel solutions to avoid liability that have not be an issue before shadow trading came to light.

[222] *Supra* note 217.

Conclusion

Insider trading is a vast field that has posed difficult to regulate. It continues to expand over time as the world evolves, businesses changes, and new legislation is enacted. Many theories of insider trading have emerged and proven to be quite successful in court. The classical theory was the initial step: this theory provided that corporate insiders trading in their own company with material and nonpublic information were guilty of insider trading. This theory worked well for a long time, but still did not address other problematic situations where securities were being purchased and sold questionably. This eventually led to the tipper/tippee theory. This theory involved situations where an insider relays information to an outsider while knowingly in breach of a fiduciary duty of trust and confidentiality. Finally, the misappropriation theory emerged which defined insider trading as any sale or purchase in the securities market where she violates any kind of duty to any person with information that was meant to be private and confidential. This served as a "catch all" system where questionable activity that did not meet the threshold of the other approaches could still be targeted for liability.

Section 10(b) of the Exchange Act sought to target any form of trading done through fraud or deceit. Over the years, this definition has been read quite broadly in order for the SEC to maintain an honest and fair securities market.

Today, the court will face yet another possible expansion of insider trading to a realm that has not previously been addressed before. *Panuwat* could potentially extend liability to those trading on similar stocks that are not directly related to a target company. *Panuwat* would push the bounds of what it means to be considered an "outsider" and could ultimately create more liability to catch activity that previously was not considered. Following precedent cases that have successfully utilized the misappropriation theory, it is likely that shadow trading will prove to be

a legal avenue for the SEC to enforce stricter insider trading rules. Courts generally allow the SEC to take an expansive and broad approach to pursuing insider trading cases. Additionally, Congress enacted legislation in an effort to maintain the integrity of the securities market. Congress left the rules broad to allow the SEC flexibility in pursuing insider trading cases. It logically follows that this will expand to also include shadow trading.

In response to potential shifts in insider trading liability, it would be wise for firms and corporations to advise their clients and employees how to avoid possible "shadow trading" situations. This could include specific language in employment contracts that eliminate confusion on what stocks can be traded on and what have to be avoided. Additionally, detailed orientations for new employees to stress company policy on securities trades and basic training on SEC regulations that are pertinent. Current client memos can provide a lot of initial guidance on how to proceed in the future.

Courts seem to further their definition of insider trading as more loopholes become apparent. The SEC is quite rigorous with insider trading specifically. It will be interesting to see how the courts will respond to "shadow trading" being the next frontier of future misappropriation theory cases. It is likely that a court will adopt this new form of insider trading and find it to be a legal route for the SEC to catch more defendants. Until then, the future of insider trading is up in the air.

NEW ASSET, SAME CUSTODY REQUIREMENTS: WHY THE EXISTING FRAMEWORK OF THE INVESTMENT COMPANY ACT AND INVESTMENT ADVISERS ACT CAN INTEGRATE DIGITAL ASSETS

By: Nicholas J. Basista, J.D.[*]

ABSTRACT

The SEC has repeatedly denied approval of ETFs that would directly own Bitcoin due to concerns of compliance with the custody requirements of the federal securities laws. The Investment Company Act and Investment Advisers Act regulate the custody of assets under the management of investment companies and advisors. The Acts, which were enacted over 80 years ago, were designed for legacy assets such as stocks or bonds, not for digital assets that exist entirely on a network of computers. While there are regulatory questions concerning digital assets in different contexts, the custody issue is perhaps the most complex. The SEC acknowledged that it is challenging for market participants to custody digital assets and that the application of the federal securities law to digital assets raises novel and complex questions and challenges.

To comply with the Acts, investment companies, advisors, and custodians must answer the following questions. How can these participants verify their holdings of digital asset? Are digital assets auditable? Can digital assets be properly segregated and accounted for? Is it possible for participants to store and secure digital assets in accordance with the Acts? Can the existing regulatory requirements be met and effectively protect investors?

This article explores the unique custody challenges participants face with respect to digital assets and argues that technology and processes exist that allow for mutual funds and ETFs to effectively meet the existing custody scheme of the Acts.

[*] Juris Doctor, Rutgers University School of Law-Camden, 2023; B.B.A., Temple University, 2017. Many thanks to Professor Arthur Laby for his help in making this article possible. I would also like to thank Professor Jay Feinman for his assistance. Lastly, I would like to thank my Mom and Dad for their encouragement over the years.

INTRODUCTION

Mutual funds and exchange-traded funds (ETFs) invest in a universe of different assets from stocks, bonds, commodities, and currencies.[1] With the launch of the first futures-based Bitcoin ETF in October 2021,[2] digital assets have entered this universe. The rise in the adoption of digital assets leaves regulators scratching their heads. The two federal securities laws that regulate the custody requirements of investment companies and their advisors were enacted over 80 years ago,[3] a time when Congress and regulators could not imagine the introduction of digital assets.

Congress's intent when passing the Investment Company Act of 1940 (ICA) and its companion statute, the Investment Advisers Act of 1940 (IAA) was to protect investors from persons controlling investment companies.[4] Prior to the passage of the ICA, the Securities and Exchange Commission (SEC) was directed to conduct a study and report on trusts and companies.[5] This study gave the SEC authority and momentum to write and propose to Congress the ICA.[6] The study revealed that persons controlling investment companies comingled client's assets and then

[1] *See* Kevin Voigt, *Exchange-Traded Fund (ETF)*, NERDWALLET (Jan. 31, 2022), https://www.nerdwallet.com/article/investing/what-is-an-etf; *see Mutual Funds*, U.S. SEC. & EXCH. COMM'N, https://www.investor.gov/introduction-investing/investing-basics/investment-products/mutual-funds-and-exchange-traded-1#:~:text=A%20mutual%20fund%20is%20a,%2C%20and%20short%2Dterm%20debt (last visited Jan. 30, 2022).

[2] Will Ashworth, *15 Bitcoin ETFs and Cryptocurrency Funds You Should Know,* KIPLINGER (Nov. 22, 2021), https://www.kiplinger.com/investing/cryptocurrency/603600/bitcoin-etfs-cryptocurrency-funds. Available crypto-backed investment company products do not directly invest in and hold digital assets; *see id.*

[3] *See The Laws That Govern the Securities Industry*, U.S. SEC. & EXCH. COMM'N, https://www.investor.gov/introduction-investing/investing-basics/role-sec/laws-govern-securities-industry (last visited Jan. 30, 2022) [hereinafter *Securities Industry Laws*].

[4] 1 TAMAR FRANKEL & ARTHUR B. LABY, THE REGULATION OF MONEY MANAGERS: MUTUAL FUNDS AND ADVISORS § 1.02, at 55 (Ann Taylor Schwing ed., 3d ed. 2021).

[5] *Id.*

[6] *See* John Morley, *Collective Branding and the Origins of Investment Fund Regulation,* 6 VA. L. & BUS. REV. 341, 347–48 (2012).

took those assets on loan.[7] Congress concluded that federal regulation was needed for those who manage the funds of investment companies, called investment advisors, resulting in the passage of the IAA.[8] To protect the safety of investor's funds, the ICA and rules accompanying the IAA prescribe conditions in which investment companies and their advisors must custody assets under their management.[9]

While there are regulatory challenges with respect to cryptocurrencies in different contexts, the custody issue raises a myriad of questions and is perhaps the most complex.[10] The nature and intangible aspect of digital assets are distinct from legacy assets such as stocks or bonds which can be certificated.[11] Unlike traditional assets which are recorded through central intermediaries in private ledgers, digital assets are recorded on decentralized digital ledgers (network).[12] Cryptocurrencies appear as entries on the network, with anonymous pseudonyms.[13] Questions arise such as how does an investment company and its related custodian verify the digital assets? How do investment companies, advisors, and other market participants store and secure the digital assets?[14] What technology is available to effectively custody the assets, and what arrangements will be sufficient to meet the regulatory requirements?[15] Can the existing regulatory requirements

[7] *See* 2 Frankel & Laby, *supra* note 4, §17.02, at 5, n.18. Before 1940, investment advisors regularly arranged for one client to buy certain securities and another client to sell the same security. *Id.* Advisors would commonly receive a proportion of the profits clients made from these transactions. *See Id.* Further, investment advisors would acquire for themselves high quality securities and place its unwanted securities in the client's account. *Id.*

[8] 1 *id.* at § 1.02, 36–37.

[9] *See* Jay G. Baris, *The Custody of Digital Assets,* GLOB. LEGAL INSIGHTS: BLOCKCHAIN & CRYPTOCURRENCY REGUL. 47, 48 (2019), https://www.shearman.com/-/media/Files/Perspectives/2018/09/GLIBLCH1ShearmanSterling-2.pdf; *see also Custody Rule Refresher: Review of Most Recent SEC Guidance,* THE NAT'L L. REV. (Nov. 20, 2018), https://www.natlawreview.com/article/custody-rule-refresher-review-most-recent-sec-guidance#:~:text=To%20comply%20with%20the%20Custody,that%20the%20other%20custodians%20provide.

[10] *See* Baris, *supra* note 9, at 47.

[11] *Id.* at 55–56.

[12] CONG. RSCH. SERV., R46208, DIGITAL ASSETS AND SEC REGULATION, 1 (2021).

[13] *See* Baris, *supra* note 9, at 56.

[14] *See id.* at 47.

[15] *See id.*

effectively protect investors? As we can see, the custody requirements of the ICA and IAA were not designed to apply to digital assets. This note explores the unique challenges funds, advisors, and custodians face in meeting the custody requirements of the Acts and argues that the existing custody framework can integrate digital assets.

Part I defines investment companies and provides a primer on digital assets. An explanation of investor demand for crypto- backed investment company products is given, along with an overview of the current state of crypto-backed investment company products. Part II explains the custody requirements of the ICA and IAA. Part III explores the challenges investment companies and associated parties face with directly holding digital assets and complying with the custody requirements. Part IV explains how investment companies, investment advisors, and custodians can comply with the existing custody framework of the Acts.

I. INVESTMENT COMPANIES, ADVISORS, & DIGITAL ASSETS

a. Investment Companies and Investment Advisors

Investment companies are institutional intermediaries that both issue securities and invest in securities.[16] There are three basic types of investment companies under the federal securities laws: mutual funds, closed-end funds, and unit investment trusts.[17]

Mutual funds are open-end investment companies that continuously pool money from investors and in turn, invest that money into securities.[18] Investors purchase shares directly from

[16] *See Investment Companies,* U.S. SEC. & EXCH. COMM'N, https://www.sec.gov/fast-answers/answersmfinvcohtm.html (last modified July 9, 2013).

[17] *See id*. (explaining that mutual funds are legally known as open-end companies; closed-end funds are legally known as closed-end companies; unit investment trusts are also known as UITs).

[18] *Mutual Funds*, U.S. SEC. & EXCH. COMM'N, https://www.investor.gov/introduction-investing/investing-basics/glossary/mutual-funds (last visited Jan. 7, 2022) [hereinafter *Mutual Funds*]; *see* Brian O'Connell & John

the mutual fund, and cannot purchase shares on the stock market, such as the Nasdaq Stock Exchange.[19] Investors pay the price for the mutual fund shares, which is the net asset value (NAV) per share plus fees associated with the fund.[20] The SEC regulates mutual funds, and these funds are registered with the Commission.[21] Separate entities, investment advisors, typically manage the portfolio of the mutual funds and are also registered with the SEC.[22]

Most often, a mutual fund is structured as a corporation or a business trust.[23] The owners of a mutual fund are its shareholders.[24] The mutual fund does not have employees; instead, virtually all of its functions are externally managed.[25] The fund's operations are administered through affiliated organizations and contractors.[26] An investment advisor manages the fund's portfolio in accordance with the objectives outlined in the mutual fund's prospectus.[27] The investment advisor or "fund manager" is responsible for making all final investment decisions, determining a fund's investment goals, and monitoring the performance and investment activity of the fund.[28] A custodian is employed and is responsible for maintaining and securing the fund's

Schmidt, *What are Closed-End Funds? How Do They Work?,* FORBES, https://www.forbes.com/advisor/investing/closed-end-funds/ (last updated Apr. 16, 2021) (ETFs are also structured as an open-end investment companies).

[19] *See Mutual Funds*, *supra* note 18.

[20] *Id.*

[21] *Id.*

[22] *Id.*

[23] *A Guide to Understanding Mutual Funds*, INV. CO. INST. 16 https://www.ici.org/system/files/attachments/pdf/g2understanding.pdf (last visited Feb. 8, 2020) [hereinafter *A Guide to Mutual Funds*].

[24] *See id.* at 5.

[25] *Id.* at 16.

[26] *Id.*

[27] *Id.*

[28] *Role of Investment Adviser and Sub-advisor*, FS INV. SOL., https://fsinvestments.com/education/role-of-investment-adviser-and-sub-adviser/#:~:text=An%20investment%20adviser%2C%20or%20%E2%80%9Cfund,to%20the%20fund's%20investment%20objectives.&text=Monitor%20and%20oversee%20the%20performance,the%20fund%20and%20sub%2Dadvis%20r (last visited Jan. 31, 2022).

assets.[29] Other affiliated organizations include a transfer agent, an independent public accountant, and a distributor.[30]

While ETFs are not mutual funds, these products combine features of mutual funds and allow investors to pool their money into the fund which reinvests the money into securities or assets.[31] Where mutual funds directly sell shares to investors, ETF shares are bought and sold on stock exchanges at a market price that is correlated to the NAV of the shares.[32] Similar to mutual funds, ETFs are professionally managed through investment advisors who are required to register with the SEC.[33]

Unlike open-end funds which continuously pool money from investors,[34] closed-end funds invest the money raised in an IPO into its portfolio.[35] Funds of the closed-end type generally do not continuously offer their shares but sell a fixed number of shares at one time.[36] Like open-end funds, closed-end funds are subject to SEC regulation and must register with the Commission.[37] The investment portfolios are also managed by investment advisors who are registered with the Commission.[38]

The last type of investment company, the unit investment trust (UITs), invests money raised from investors in a one-time public offering in a portfolio that is generally fixed.[39] UITs do

[29] *See A Guide to Mutual Funds, supra* note 23.
[30] *See id.*
[31] *See Mutual Funds and Exchange-Traded Funds (ETFs) – A Guide for Investors,* U.S. SEC. & EXCH. COMM'N, https://www.sec.gov/reportspubs/investor-publications/investorpubsinwsmfhtm.html#ETFs (last modified Jan. 26, 2017).
[32] *Id.*
[33] *See id.*
[34] *Mutual Funds, supra* note 18.
[35] *Closed-end Funds*, U.S. SEC. & EXCH. COMM'N, https://www.investor.gov/introduction-investing/investing-basics/glossary/closed-end-funds (last visited Jan. 7, 2022).
[36] *Id.*
[37] *See id.*
[38] *Id.*
[39] *Unit Investment Trusts (UITs)*, U.S. SEC. & EXCH. COMM'N, https://www.investor.gov/introduction-investing/investing-basics/glossary/closed-end-funds (last visited Jan. 7, 2022).

not actively trade securities in their portfolio; instead, it buys a certain fixed amount of securities and holds them for the life of the trust.[40] Unlike the other two types of investment companies, a UIT does not have an advisor to manage its portfolio; however, the UIT must register with the SEC and is subject to SEC regulation.[41]

b. Understanding Digital Assets

The SEC has stated that "'digital assets' include instruments that *may* qualify under applicable U.S. securities laws . . ."[42] The existence and transfer of digital assets, such as Bitcoin are recorded on a distributed ledger across a network of decentralized computers that validate the transactions.[43] The design of the network is intended to create trust amongst users as there is no centralized authority such as a government.[44] The decentralized computer network relies heavily on cryptography to ensure security.[45] When transacting in cryptocurrency, the price to purchase or sell a unit is usually quoted in government-based fiat currency.[46] The majority of cryptocurrency networks allow for the transfer of the asset between users, with no need for a centralized party and without identification between users.[47]

Distributed ledger technology, which is a form of blockchain technology, is at the heart of cryptocurrencies.[48] The underlying technology allows parties to access the network and its

[40] *Id.*
[41] *Id.*
[42] *Leaders of CFTC, FinCEN, and SEC Issue Joint Statement on Activities Involving Digital Assets*, U.S. SEC. & EXCH. COMM'N, (Oct. 11, 2019), https://www.sec.gov/news/public-statement/cftc-fincen-secjointstatementdigitalassets (emphasis added).
[43] *See* Gabriel Abed et al., *Navigating Cryptocurrency Regulation*, WORLD ECON. F.: GLOB. FUTURE COUNCIL ON CRYPTOCURRENCIES 1 (Sept. 2021), https://www3.weforum.org/docs/WEF_Navigating_Cryptocurrency_Regulation_2021.pdf.
[44] *Id.* at 3.
[45] *See id*. at 5.
[46] *See id.*
[47] *See id.*
[48] Marco Iansiti & Karim R. Lakhani, *The Truth About Blockchain,* HARV. BUS. REV. (Jan – Feb. 2017), https://hbr.org/2017/01/the-truth-about-blockchain.

complete history.[49] The decentralized aspect of the network prohibits a single party from controlling the data.[50] Once transactions are validated, and accounts are updated, it is impossible to alter the ledger.[51] For purposes of this note, the terms "cryptocurrency," "crypto assets," and "digital assets" are used interchangeably.

As we will see below, how investors access and transact in cryptocurrency is important. Major cryptocurrencies such as Bitcoin, are built on public key cryptography, which utilizes public keys and private keys.[52] Public keys are cryptographic code that display to all users who access the ledger identifying information of the other users.[53] While a private key is a string of letters and numbers picked at random, "ownership and control over the private key is the root of user control over all funds associated with the corresponding [digital asset] address."[54] Private keys are kept confidential and are necessary for authentication and effectuating transactions.[55] There are two main ways to store a private key. In what is known as a "hot" wallet (hot storage), where the

[49] *Id.*

[50] *Id.*

[51] *Id.*

[52] *Public and Private Keys,* BLOCKCHAIN SUPPORT, https://support.blockchain.com/hc/en-us/articles/360000951966-Public-and-private-keys (last visited Jan. 21, 2022) [hereinafter *Public and Private Keys*]; *see* also Amanda Gould, *Crypto Custody*, Penn L., https://www.law.upenn.edu/faculty/david-hoffman/crypto-custody.php#_edn80 (last visited Feb. 1, 2022) (providing that the public and private keys are stored on what is called a wallet).

[53] *Id.; see also* Jake Frankenfield, *Cold Storage*, Investopedia, https://www.investopedia.com/terms/c/cold-storage.asp#:~:text=Cold%20storage%20is%20a%20way,their%20holdings%20via%20traditional%20means (last updated June 30, 2020) ("The public key is akin to an account name and helps to identify a destination for [digital assets] that are being sent to [a] wallet.").

[54] ANDREAS M. ANTONOPOULOS, MASTERING BITCOIN 89 (O'Reilly Media, ed., 2d ed. 2017); *see also* Ryan Haar, *How to Decide on a Hot Wallet or Cold Wallet for Your Crypto, and Whether You Need One at All,* Next Advisor (Sept. 23, 2021), https://time.com/nextadvisor/investing/cryptocurrency/hot-wallet-vs-cold-wallet/ (providing that the private key is akin to a bank account password, which allows access to the wallet and to purchase or sell the digital asset).

[55] *Public and Private Keys*, *supra* note 52.

associated address is stored on the internet.[56] Or the user can opt for "cold storage" where the associated address is kept offline, such as written on a piece of paper or stored on a flash drive.[57]

c. Demand for Crypto-Backed Investment Company Products

Investing in digital assets, such as Bitcoin can be complicated for investors.[58] For retail investors, investing in this asset class requires extra knowledge and work in comparison to investing in stocks, bonds, and similar assets.[59] Crypto ETFs and mutual funds allow investors to gain exposure to cryptocurrencies more easily.[60] Physically holding bitcoin can be tricky to secure and store.[61] Investors may forget their passwords needed to access their digital assets.[62] For investors who directly purchase bitcoin, their transactions are stored in a wallet, which is unlocked with the private key.[63] A study from Chainanalysis, a cryptocurrency research firm, shows that about 18% of Bitcoin is stuck in inaccessible wallets or lost due to users forgetting or misplacing the private keys.[64] Another option for investors is to invest in crypto assets through exchanges.[65] Crypto exchanges require more effort in comparison to traditional brokerage accounts.[66] Even though mainstream crypto exchanges have made purchasing digital assets more accessible, ETFs

[56] *Id.* Hot wallets allow for easy access to the asset to facilitate trades; however, because of the internet connection, these wallets are not secure from hacks. *See id.*

[57] *Id.* Cold storage allows for greater security against hacks; however, this form of storage makes it harder to access digital assets. *See id.*

[58] *See* Alana Benson & Kurt Woock, *Bitcoin ETFs: What to Know,* NERDWALLET (Dec. 21, 2021), https://www.nerdwallet.com/article/investing/bitcoin-etfs.

[59] *See* Nicholas Rossolillo, *Investing in Bitcoin ETFs,* THE MOTLEY FOOL (Jan. 18, 2022, 12:39 PM), https://www.fool.com/investing/stock-market/market-sectors/financials/cryptocurrency-stocks/bitcoin-etfs/.

[60] Benson et al., *supra* note 58.

[61] *Id.*

[62] *Id.*

[63] *Id.*

[64] Benson et al., *supra* note 58.

[65] *What You Need to Know About Crypto Mutual Funds,* EARLY BIRD (Dec. 27, 2021), https://www.getearlybird.io/blog/crypto-mutual-fund [hereinafter EARLY BIRD].

[66] *See id*. Investors are required to set up a new account with a trusted crypto exchange and become familiar with crypto safety standards. *Id.*

can be bought and sold from investor's existing brokerage accounts.[67] Mutual funds and ETFs with exposure to cryptocurrency offer tax-advantaged investing, and mutual funds and ETFs are easy to buy in retirement accounts.[68]

d. State of Crypto Backed Investment Products

There are currently several crypto asset linked ETFs available to investors.[69] The first ETF to launch was the ProShares Strategy ETF (trading as BITO) which launched in October of 2021.[70] Prior to SEC approval of BITO, there were up to 13 applications awaiting SEC approval.[71] Currently, there are not any available ETFs or mutual funds that are able to directly own Bitcoin.[72] In November 2021, the SEC denied the approval of VanEck ETF, which would directly hold Bitcoin.[73] Available Bitcoin ETFs such as BITO, "[do]not invest directly in Bitcoin, which provides as close to one-to-one exposure as you could want."[74] BITO instead invests in cash settled futures contracts.[75] "SEC Chair Gary Gensler has said in the past that he would prefer to see funds holding Bitcoin futures rather than the cryptocurrency itself."[76] The focus of this note is solely on investment companies directly holding digital assets.

---

[67] *See* Benson et al., *supra* note 58; *see also* Rossolillo, *supra* note 59 (explaining that traditional brokerage firms do not allow cryptocurrency trading, investors need to open an account with a crypto trading exchange).

[68] EARLY BIRD, *supra* note 65.

[69] *See* Ashworth, *supra* note 2.

[70] *Id.*

[71] *Id.*

[72] Rossolillo, *supra* note 59; *see also* Ashworth, *supra* note 2 (explaining these types of funds are known as spot funds).

[73] Ashworth, *supra* note 2.

[74] Id.; see also Taylor Locke, *Bitcoin Futures ETF Started Trading Today, But One Crypto Expert Says It's "Not Something Retail Investors to Buy,"* CNBC (Oct. 19, 2021, 11:33 AM) (explaining that a Bitcoin futures ETF tracks contracts that speculate on the future price, not the "spot price" of the asset itself, which means the price of the futures linked ETF could trade at a premium when the market is high or vice versa).

[75] *Id.;* Prableen Bajpai, *How to Invest in Bitcoin Futures,* INVESTOPEDIA, https://www.investopedia.com/articles/investing/012215/how-invest-bitcoin-exchange-futures.asp#:~:text=Bitcoin%20futures%20enable%20investors%20to,on%20the%20cryptocurrency's%20future%20price (last updated Dec. 3, 2021) ("Bitcoin futures enable investors to gain exposure to Bitcoin . . . without having to hold the underlying cryptocurrency.").

[76] Ashworth, *supra* note 2.

II. CUSTODY REQUIREMENTS OF THE ICA AND IAA

a. Custody Requirements Under the ICA

The ICA is the principal statute regulating the organization of investment companies, such as mutual funds and ETFs.[77] For our purposes, we will examine the SEC enforced custody obligations the ICA imposes on investment companies concerning custody of their holdings and the holding of their assets with a broker dealer.[78] Primarily, this is Section 17(f) of the ICA and an accompanying Rule.[79] Section 17(f) prescribes how investment companies must custody their assets.[80] The ICA defines assets that fall under its purview as "securities *and similar investments*," which can be readily read to cover digital assets.[81] An investment company must place and secure its securities and similar investments in the custody of either: (1) a bank meeting certain requirements; (2) member of a national securities exchange (i.e., a broker-dealer), subject to any SEC rules; or (3) the company itself.[82]

The accompanying Rule specifies the requirements where an investment company can use a broker-dealer as a custodian.[83] The rule requires a written contract between the investment

[77] *See Laws and Rules*, U.S. Sec. & Exch., https://www.sec.gov/investment/laws-and-rules (last visited Jan. 23, 2022).

[78] James Chen, *Investment Company Act of 1940,* INVESTOPEDIA, https://www.investopedia.com/terms/i/investmentcompanyact.asp#:~:text=Key%20Takeaways-,The%20Investment%20Company%20Act%20of%201940%20is%20an%20act%20of,and%20Exchange%20Commission%20(SEC) (last updated May 19, 2021) ("The legislation in the [ICA] is enforced and regulated by the Securities and Exchange Commission . . .").

[79] Investment Company Act of 1940, 15 U.S.C. § 80a-17(f) (2018); 17 C.F.R. § 270.17f-1(a) (2022); *see also* Custody of Investment Company Assets with a Securities Depository, Investment Company Act Release No. IC-25934 (Feb. 13, 2003).

[80] 15 U.S.C. § 80a-17(f); 17 C.F.R § 270.17f-1(a).

[81] 15 U.S.C. § 80a-17(f); 17 C.F.R. § 270.17f-1(a); s*ee* Gould, *supra* note 52 ("A registered fund investing solely or partially in crypto assets deemed to be securities will almost certainly invoke the '40 Act custody provisions.").

[82] 15 U.S.C. § 80a-17(f)(1); 17 C.F.R. § 270.17f-1(a).

[83] 15 U.S.C. § 80a-17(f)(1); 17 C.F.R. § 270.17f-1(b)(1).

company and the broker dealer.[84] The contract must specify that the securities of the investment company be physically segregated from the securities of other persons and marked clearly to identify the client's (i.e., the investment company's) securities.[85] Additionally, the securities in the custody of the broker-dealer must be verified by actual examination each annual and semiannual year through an independent public accountant and subject to examination from the SEC.[86]

b. Custody Requirements Under the IAA

The IAA initially did not include a provision governing the custody of funds under the management of advisors.[87] Instead, the SEC later adopted a rule governing custody of investment advisor's assets.[88] For our purposes, we will focus on this SEC enforced custody rule.[89]

Rule 206(4)-2 (the Custody Rule) under the IAA applies to advisors registered with the SEC and to custodians who maintain the assets.[90] The Custody Rule defines custody as "holding, directly or indirectly, client funds or securities, or having authority to obtain possession of them."[91]

The Custody Rule requires either a bank, broker-dealer, or a futures commissions merchant to custody the assets of the advisor.[92] Under the Custody Rule, the custodian must segregate client

[84] 17 C.F.R. § 270.17f-1(a).
[85] 17 C.F.R. § 270.17f-1(b)(1).
[86] 17 C.F.R. § 270.17f-1(b)(4)–(5).
[87] *See* Kelley A. Howes, *The Custody Rule Under the Investment Advisers Act: Time for a Change*, IAA NEWSL. (Nov. 2017), https://media2.mofo.com/documents/171100-custody-rule-investment-change.pdf.
[88] *Id.* In 1962, the SEC adopted Rule 206(4)-2 "which for the first time, required registered investment advisers to implement a set of controls to protect client assets over which they maintained custody." *Id.*
[89] *General Information on the Regulation of Investment Advisers,* U.S. SEC. & EXCH., https://www.sec.gov/divisions/investment/iaregulation/memoia.htm#:~:text=The%20Securities%20and%20Exchange%20Commission,(the%20%22rules%22) (last updated Mar. 31, 2017). The SEC, "regulates investment advisers, primarily under the [IAA], and the rules adopted under that statute." *Id.*
[90] 17 C.F.R. § 275.206(4)-2(a)(1) (2022).
[91] 17 C.F.R. § 275.206(4)-2(d)(2).
[92] 17 C.F.R. § 275.206(4)-2(6)(i)-(iv).

funds for each client under that client's name and cannot comingle the assets.[93] In addition, there is a surprise audit requirement and an account statement requirement under the Custody Rule.[94]

The surprise audit requirement mandates that at a minimum, once a year, the advisor and the qualified custodian have the custodied assets verified in a surprise audit where the custodied assets are accounted for.[95] An independent public accountant conducts the audit, and the date is chosen by the accountant with no notice given to either the advisor or the qualified custodian.[96]

The account statement requirement mandates that the custodian or advisor send account statements to each client who has invested in the given fund.[97] These statements must be sent quarterly.[98] The statements must specify the amount of assets in the account of each client at the end of each period.[99] The statements must also include all transactions that occurred during the period.[100]

III. CHALLENGES FACED WITH CUSTODY OF DIGITAL ASSETS

As discussed above, investment companies and their advisors can comply with the requirements of the ICA and IAA through delegating the responsibility of custody to third-party qualified custodians, such as broker-dealers. While this solves half the problem, the question remains, how does the qualified custodian comply with the custody requirements?[101] In what manner can the custodian custody the digital assets that satisfies regulatory scrutiny and adequately

[93] *See* 17 C.F.R. § 275.206(4)-2(a)(i)(ii).
[94] *See* 17 C.F.R. § 275.206(4)-2(a)(4).
[95] *Id.*
[96] *Id.*
[97] 17 C.F.R. § 275.206(4)-2(a)(3).
[98] *Id.*
[99] *Id.*
[100] *Id.*
[101] Baris, *supra* note 9, at 52.

protects the assets?[102] How much protection is there to protect against fraud and malicious third parties? What makes digital assets unique from legacy securities is that they exist entirely and are verifiable only by nodes on a digital ledger.[103] So, in what way can independent auditors verify the assets? Can the assets be properly segregated? Can the advisor send client statements of the transactions? The SEC and other regulatory agencies ask these same questions and more.

In a joint statement, the staff of the SEC and the Financial Industry Regulatory Authority (FINRA) acknowledged that it is challenging for market participants to custody digital assets.[104] The staffs noted, "[a]s a threshold matter, it should be recognized by market participants that the application of the federal securities laws . . . [to] digital asset securities . . . raise novel and complex regulatory questions and challenges."[105]

In February 2021, the SEC issued a Risk Alert concerning digital assets.[106] The Alert identified risks of investment advisors managing digital assets.[107] Specifically, the Alert stated the SEC "[would] review the risks and practices related to the custody of digital assets by investment advisors and examine for compliance with the [Custody Rule]."[108] The staff noted it would review the frequency of unauthorized transactions, theft of digital assets, and the mechanisms and controls for safekeeping the private keys.[109] The staff would examine business continuity plans for key

[102] *Id.*

[103] *Id.* at 55.

[104] Div. of Trading and Mkts., U.S. Sec. and Exch. Comm'n & Off. of Gen, Couns., Fin. Indus. Regul. Auth., *Joint Staff Statement on Broker-Dealer Custody of Digital Asset Securities,* U.S. SEC. AND EXCH. COMM'N (July 8, 2019), https://www.sec.gov/news/public-statement/joint-staff-statement-broker-dealer-custody-digital-asset-securities.

[105] *Id.*

[106] *The Division of Examinations' Continued Focus on Digital Asset Securities,* U.S. SEC. AND EXCH. COMM'N 1 (Feb. 26, 2021), https://www.sec.gov/files/digital-assets-risk-alert.pdf.

[107] *Id.*

[108] *Id.* at 3.

[109] *Id.*

personnel who have access to private keys.[110] The SEC would review the security procedures related to software and hardware wallets.[111] In addition to addressing the risks associated with investment advisors, the Risk Alert identified concerns regarding the custody of digital assets with broker-dealers.[112] The staff's examination centered on broker-dealer's operational activities, and operations that are unique to the safety and security regarding custody of digital assets.[113]

In December 2020, the SEC issued a statement and request for comment concerning broker-dealers custodying digital assets.[114] Similar to concerns in the Risk Alert, the SEC requested comment on the protection of private keys and best practices against theft, loss, generation, and unauthorized use of private keys.[115] The Commission asked for industry best practices for using private keys.[116] The SEC also asked for a response to evaluate events such as a hard fork, airdrop, and 51% attack"[117]

This SEC statement provides some guidance for broker-dealers seeking to custody digital assets. The statement, however, does not make clear what is required under the IAA or ICA. Nor does the statement give guidance on whether broker-dealers may act as qualified custodians for purposes of the ICA or IAA. The SEC noted it "will continue to evaluate its position, and the circumstances . . . on an ongoing basis . . . ."[118]

---

[110] *Id.*
[111] *Id.*
[112] *Id.* at 3 – 4.
[113] *Id.* Other concerns noted in the Risk Alert regarding broker-dealers were anti-money laundering concerns, exchange registration, and compliance with Regulation ATS. *Id.*
[114] Press Release, SEC Issues Statement and Requests Comment Regarding the Custody of Digital Asset Securities by Special Purpose Broker-Dealers (Dec. 23, 2020), https://www.sec.gov/news/press-release/2020-340.
[115] Custody of Digital Asset Securities by Special Purpose Broker-Dealer, Exchange Act Release No. 34-90788 (Dec. 23, 2020), https://www.sec.gov/rules/policy/2020/34-90788.pdf [hereinafter SEC SPDB Release].
[116] *Id.*
[117] *Id.*
[118] *Id.*

While not explicitly noted by the SEC, digital assets make it difficult for independent auditors to verify the assets and for advisors to properly segregate the client's assets. With legacy assets such as stocks, a custodian can take physical possession of a stock certificate.[119] For an uncertificated stock certificate, the custodian can take possession as the custodian's name is registered "on the books of the issuer."[120] This is not the case for digital assets. Crypto assets are considered bearer assets.[121] In other words, whoever has access to the private key is considered the owner.[122] Digital asset transactions are reflected in a distributed ledger, which reflects the transfer of ownership, but this does not always indicate who is the owner of the asset.[123]

IV. WHY DIGITAL ASSETS FIT INTO THE EXISTING FRAMEWORK

The statements from the SEC indicate its belief digital assets do not fit into the custody framework of the Acts. The statements also telegraph the SEC's position that digital assets pose risks too great for market participants to contain. While the SEC's concerns are well-founded, institutions and market participants are adapting to and creating a technological ecosystem around digital assets. The advancements in technology allow for investment companies, investment advisors, and custodians to comply with the custody requirements of the ICA and IAA, in addition to ensuring investors' assets are appropriately safeguarded.

---

[119] Baris, *supra* note 9, at 56.

[120] *Id.*

[121] *Can You Prove That You Own Your Crypto Assets?, TRANSITNET,* https://transitnet.io/blog/can-you-prove-that-you-own-your-crypto-assets/ (last visited Jan. 22, 2022) (explaining that a bearer asset is an asset where no identifying or ownership information is recorded and associated with the asset).

[122] *Id.*

[123] Betsy Vereckey, *Bitcoin: Who Owns It, Who Mines It, Who's Breaking the Law*, MIT MGMT, https://mitsloan.mit.edu/ideas-made-to-matter/bitcoin-who-owns-it-who-mines-it-whos-breaking-law (Oct. 14, 2021) (explaining with bitcoin, transactions of the asset contain a list of senders and recipients represented by an anonymous address, the number of bitcoins sent and received, and a timestamp of when the transaction occurred).

a. Custody of Digital Assets and Storage of Private Keys

The SEC has identified the issue of best practices for management of private keys[124] and protection against theft, loss, and unauthorized or accidental use of private keys.[125] The SEC's concern regarding private keys is well placed.[126] In the case of a lost or stolen private key, the funds associated with that key are lost.[127] There is also the concern of a malicious individual within the organization using the private key for unauthorized transactions. Lastly, there is the concern that hackers or other malicious third parties may gain access to the private key. Custodians have several options to safeguard the private key, which include multi-signature wallets, multi-party computation, and single-purpose hardware security module wallets.

[124] Jake Frankenfield, *Private Key,* Investopedia, https://www.investopedia.com/terms/p/private-key.asp#:~:text=A%20cryptocurrency%20wallet%20consists%20of,control%20and%20ownership%20of%20cryptocurrency (last updated Nov. 27, 2021) (It is important to reiterate, "[p]rivate keys represent final control and ownership of cryptocurrency.").

[125] *See supra* text accompanying notes 108 – 11, 115 – 16.

[126] *See* Colin Harper, *Multisignature Wallets Can Keep Your Coins Safer (If You Use Them Right),* CoinDesk (Nov 10, 2020, 4:17 PM), https://www.coindesk.com/tech/2020/11/10/multisignature-wallets-can-keep-your-coins-safer-if-you-use-them-right/ (explaining that companies engaging in digital assets are prone to a literal version of key person risk when handling digital assets) ("The most infamous example may be QuadrigaCX, whose customers have been waiting nearly three years to recoup $115 million worth of deposits since the death of founder Gerald Cotton, the sole possessor of the cryptographic keys to the exchange's wallet.").

[127] *See id.*

Multi-signature wallets (multisig) are an effective solution to safeguard digital assets[128] and protect against theft, loss,[129] and unauthorized use of private keys.[130] The standard of multisig wallets has been embraced by many companies and organizations since their inception in 2013.[131]

Multisig wallets can be analogized to a bank vault that requires multiple keys to open because these wallets require at least two private keys to access the wallet and effectuate a transaction.[132] Multiple signing keys are given to individuals and set parameters govern the number of signing keys needed to authorize a transaction.[133] The parameters are described by "n" which represents the number of independent keys and "m" which represents the number of those keys needed to authorize a transaction.[134] Regarding custodians holding digital assets, five keys would be distributed to appropriately-authorized employees.[135] Of those five keys, two or three keys would be needed to authorize a transaction.[136] A protocol where three of the five keys are needed to authenticate a transaction, is known as a "3 of 5" system.[137] As an added measure of protection, each authorized employee would authenticate their respective key using a different hardware key

[128] *A Market Overview of Custody for Digital Assts: Digital Custodian Whitepaper,* DELOITTE 7 (June, 2020), https://www2.deloitte.com/content/dam/Deloitte/xe/Documents/finance/me_Digital-Custodian-Whitepaper.pdf ("Custodians safeguard digital assets by ensuring that investors' private keys are maintained securely . . . through a multiple approval approach, known as multi-signature . . . *".*

[129] *See* Wietze Bronkema, *Introduction to Multisig Contracts,* MEDIUM.COM (Jan. 15, 2020), https://medium.com/mycrypto/introduction-to-multisig-contracts-33d5b25134b2 (noting multisig wallets have a number of use cases including "[a]voiding a single point of failure, which makes it significantly harder for funds to be compromised.*"); see also* Brent Perreault, *Simple Multisig: How it Works and Why It's Awesome,* PAXOS: BLOG (Mar. 23, 2021), https://paxos.com/2021/03/23/simple-multisig-how-it-works-and-why-its-awesome/.

[130] *See* Perreault, *supra* note 129 (noting multisig wallets split the responsibility of the private key between multiple individuals resulting in less of a chance of a rouge personnel misappropriating the assets).

[131] Perreault, *supra* note 129.

[132] Harper, *supra* note 125.

[133] Perreault, *supra* note 129.

[134] *Id.* The parameters described is known as an "m of n wallet." *Id.*

[135] Letter from Benjamin S. Kaplan, Co-CEO, Prometheum, Inc., to Vanessa A. Countryman, Sec'y, SEC 3 (Apr. 26, 2021), https://www.sec.gov/comments/s7-25-20/s72520-8734178-237104.pdf [hereinafter Prometheum Letter].

[136] *See id.*

[137] *See id.*

system.[138] Each of the hardware systems would be under the control of and integrated through the custodian.[139]

The implementation of a multisig wallet protects against internal bad actors up to (m – 1) and against the loss of a signing key up to (n – m).[140] Using the above "3 of 5" system as an example, the protocol can withstand up to two internal malicious actors without allowing unauthorized transactions.[141] The protocol can also withstand two authorized employees losing the signing keys before the assets are lost.[142] Further, multisig wallets provide for greater security against hacks as there is no longer a single point of failure.[143] Multisig wallets solves the issue of key personnel becoming incapacitated or losing the key in another fashion.[144] "Essentially, the sum of the 'pieces' will ensure the whole is not lost."

In addition to the protections provided from multisig wallets, the custodian can implement smart contracts in conjunction with the wallets to provide a replacement and key rotation system.[145] In the case of a security breach, the keys in place within the system should be replaced with less-frequently used keys.[146] Keys which are regularly used within a protocol should also be replaced.[147] A system providing key rotation or replacement can be implemented to require all other keys used in the protocol to conform with the replacement key or the system can use a separate key which is held in a different location or with a separate entity.[148]

---

[138] *Id.*
[139] *Id.*
[140] *See* Perreault, *supra* note 129.
[141] *See id.*
[142] *See id.*; *see also supra* text accompanying notes 109 - 11.
[143] Perreault, *supra* note 129.
[144] Prometheum Letter, *supra* note 135.
[145] *See id*.
[146] *Id.*
[147] *Id.*; *see supra* text accompanying note 115.
[148] Prometheum Letter, *supra* note 135.

Another option for advisors and custodians is multi-party computation wallets (MPC). MPC may be considered the best option for private key infrastructure and very well may emerge as the industry best practice.[149] The goal of MPC technology is to provide "bank-grade protection" of private keys against misuse or theft.[150] MPC technology "is a cryptographic protocol that distributes a computation across multiple parties where no individual party can see the other parties' data."[151] Through the use of an MPC, a set number of participants (i.e., devices) are given a piece of private information, the participants bring together their respective private information to compute the function.[152] In the context of a digital asset custody wallet, the multiple devices are responsible for a portion of the private key (key shares) and those portions of the private key are brought together in a multiparty approval transaction.[153] This "joint computation" results in a signature which releases the digital assets for effectuating transactions.[154] The technology enables the custodian to "compliantly, securely, and privately compute on distributed data without every exposing or moving it."[155] MPC technology offers many benefits, it is commercially available,[156] no trusted third party is capable of accessing the data, it allows for secure operations even where parties are corrupted, and the protocol is quantum safe.[157] Further, MPC custody wallets when properly implemented allows for the enhanced security of cold storage and the accessibility to

[149] *See* Frank Wiener, *Secure Multiparty Computation (MPC) for Digital Asset Custody Wallets,* SEPIOR (Feb. 9, 2021), https://sepior.com/blog/secure-multiparty-computation-mpc-for-digital-asset-custody-wallets; *see also What is MPC (Multi-Party Computation)?,* FIREBLOCKS, https://www.fireblocks.com/what-is-mpc/. [hereinafter Fireblocks] (noting large financial institutions such as Celsius, the largest U.S. crypto lending desk, and Revoult, Europe's largest neobank, have transitioned to MPC technology).

[150] *Wiener*, *supra* note 149.

[151] *What is Secure Multiparty Computation,* INPHER, https://inpher.io/technology/what-is-secure-multiparty-computation/ (last visited Feb. 1, 2020).

[152] See *Wiener*, *supra* note 149.

[153] *Id.*

[154] *See id.*

[155] *What is Secure Multiparty Computation, supra* note 151.

[156] *Why KPMG is Mistaken About Crypto Custody*, COPPER (Mar. 31, 2020), https://copper.co/insights/why-kpmg-is-mistaken-about-crypto-custody,

[157] *See* Wiener, supra note 149.

assets offered from hot storage.[158] MPC allows for the storage of the private key online, which allows for easier access to the assets than cold storage, and allows for enhanced security because the private key is divided into many parts stored on separate devices.[159]

The implementation of MPC wallets also allows for a built-in key rotation system.[160] Because private keys are binary numbers, the different parts are a sequence of values that when jointly computed, equals the private key.[161] The sequence of the key shares used in MPC can be changed, which does not change the actual private or public key.[162] Currently, there is at least one commercially available algorithm which automatically refreshes MPC key shares in minute-long intervals.[163] Refreshing or rotating the key values, results in the reduction that a hacker or malicious third party could gain access to enough devices to derive the key.[164] With the implementation of the key refresh algorithm described above, a malicious third party will only have moments to gain access to the various devices before the key shares are refreshed.[165] This adds an additional security measure for the custodian.

The last option for advisors and custodians to use for custody and safeguarding of digital assets is single-purpose hardware security modules (HSMs).[166] HSM technology is a highly trusted, specialized device that is used for encryption, authentication, key management, and other

---

[158] *Id.; see also Copper, supra* note 156.

[159] Letter from Alan Konevsky, Chief Legal Officer, tZERO, to Vanessa A. Countryman, Sec'y, SEC 4 (Apr. 7, 2021), https://www.sec.gov/comments/s7-25-20/s72520-8648975-231028.pdf [hereinafter tZero Letter].

[160] *See* Wiener, supra note 149.

[161] *See id.*

[162] *Id.*

[163] *See Introducing MPC-CMP: Pushing Wallet Signing Speeds 8X,* FIREBLOCKS (May 13, 2020), https://www.fireblocks.com/blog/pushing-mpc-wallet-signing-speeds-8x-with-mpc-cmp-9/.

[164] *See id.*

[165] *See id.*

[166] Letter from Nathan McCauley, Co-founder & CEO, Anchorage Digit. & Georgia Quinn, Gen. Couns., Anchorage Digit., to Vanessa A. Countryman, Sec'y, SEC (July 26, 2021), https://www.sec.gov/comments/s7-25-20/s72520-9088935-246726.pdf [hereinafter Anchorage Letter].

cryptographic operations.[167] The technology is mature, regularly tested and there are existing governmental standards on which HSMs are rated.[168] Today, HSM technology is used in industries, such as banking.[169]

An HSM wallet generates and stores the private key in an encrypted tamper-proof environment.[170] HSM wallets restrict network access as this type of wallet utilizes a firewall.[171] HSMs allow for the custodian to employ strict control access, and HSMs are "virtually impossible to compromise."[172] The custodian would keep the HSM wallet offline, which would further defend against a breach.[173] With the HSM stored offline, the hacker would need to physically access the HSM.[174] HSMs allow for the generation, rotation, and protection of private keys for the custodian.[175]

While this note does not opine on which of these methods is the best option for the storage and security of digital assets, investment companies, fund managers, and custodians have at least three options for the storage of digital assets and private keys.

b. Auditing, Segregation, and Account Statements

---

[167] *What is an HSM? What are the benefits of using an HSM?,* ENCRYPTION CONSULTING, https://www.encryptionconsulting.com/education-center/what-is-an-hsm/#compliance (last visited Feb. 2, 2022) [hereinafter Encryption Consulting]; *see also What are Hardware Security Modules (HSMs)?,* ENTRUST, https://www.entrust.com/resources/hsm/faq/what-are-hardware-security-modules#:~:text=Hardware%20security%20modules%20(HSMs)%20are,creating%20digital%20signatures%20and%20certificates (last visited Jan. 31, 2022) ("HSMs are tested, validated and certified to the highest security standards.").

[168] Anchorage Letter, *supra* note 166 ("HSMs with a rating of FIPS 140-2 meet the rigorous security requirements the National Institute of Standards and Technology has laid out for cryptographic modules*.")*.

[169] Dror Trieman, *Can HSM technology make blockchain wallets and transactions safer?,* MEDIUM (Jan. 7, 2019), https://medium.com/orbs-network/can-hsm-technology-make-blockchain-wallets-and-transactions-safer-8e91dc48f1a2.

[170] *Helenix Delivers Trusted Cryptocurrency HSM Wallet Based on Entrust nShield HSM*, ENTRUST, https://www.entrust.com/-/media/documentation/solution-briefs/nshield-helenix-hsm-wallet-sb.pdf (last visited Feb. 26, 2022).

[171] *See* Encryption Consulting, *supra* note 167.

[172] *Id.*

[173] *See id.*

[174] *See id.*

[175] *Id.*

As mentioned above, digital asset networks allow for users to record and observe transfers, record the number of units transferred, amount of fiat transferred, and the time of the transaction.[176] The associated public key which appears on the network, is pseudonymized to protect privacy, meaning the records do not confirm the identity of the user.[177] With the use of technology and traditional processes, the sources provided from the network allow for auditors to track and confirm the existence and ownership of the digital assets.[178]

To ensure digital assets are appropriately segregated and to ensure auditability, the advisor and custodian would use internal controls, including a complete audit trail of the assets.[179] With the use of an audit trail, the custodian and advisor would chronologically capture events and log the movement of the specific digital asset on the network.[180] The advisor and custodian would store the specific digital asset of their single client in one digital asset custody wallet.[181] With the audit trail history and respective digital assets in a single wallet, the traditional verification practice of "books and records" would be used to verify the assets.[182] The auditor would examine supporting documents which would validate ownership of the asset.[183] The use of books and records on blockchain makes digital assets "auditable in the truest sense of the word."[184]

---

[176] Prometheum Letter, *supra* note 135, at 8; *see also* tZero Letter, *supra* note 159 ("The blockchain allows for a courtesy carbon copy of certain ownership records to be viewable on the [network].").

[177] tZero Letter, *supra* note 159.

[178] *See* Public and Private Keys, *supra* note 52.

[179] *What is a Digital Asset Management Audit Trail and Why it's so Important,* DBGALLEY, https://dbgallery.com/what_is_an_audit_trail (last visited Feb. 8, 2022).

[180] *Id.*

[181] *See supra* text accompanying notes 85, 93.

[182] *See* Nishani Edirisinghe, Vincent & Anne M. Wilkins, *Challenges When Auditing Cryptocurrencies*, 14 CURRENT ISSUES IN ACCT. 46, 54 (2019); *see also* Letter from Amy Davine Kim, Chief Pol'y Officer, Chamber of Digit. Co, to Vanessa A. Countryman, Sec'y, SEC (Apr. 5, 2021), https://www.sec.gov/comments/s7-25-20/s72520-8634796-230930.pdf [hereinafter Digital Chamber Letter] ("Accurate books and records, among other things, help ensure that investor's assets are properly custodied; allow the parties to a securities transaction to track that the transaction from order entry, to execution, and then to settlement . . .").

[183] *See* Vincent & Wilkins et al., *supra* note 182, at 55.

[184] Prometheum Letter, *supra* note 135, at 8; *see also supra* text accompanying notes 86, 95.

In addition, auditors would need to verify ownership of the private key to ensure segregation of assets and to complete the audit process.[185] The custodian would demonstrate ownership of the private key through signing a specified message from the auditor with its private key, note that this procedure does not reveal the private key nor does a transaction of digital assets or fiat occur.[186] Custodians who use HSM custody wallets, could also choose to use challenge-response authentication which "allows for objective proof of existence of crypto assets nearly instantly."[187] As a result of the auditing, tracking, and verification capabilities, the investment fund and advisor would establish procedures in the same manner as legacy assets and issue account statements in the same manner.

If the advisor uses an exchange to transact in the digital asset, auditors would be able to confirm the asset through the records with the third party exchange.[188] The auditor, advisor, and custodian would have to ensure that the third party exchange's service organization controls (SOC) are sufficient to ensure proper verification of the assets.[189] Major exchanges, such as Crypto.com and Gemini[190] are SOC 2 certified.[191] This method may prove to be another viable path for verification and auditing of the custodian's and advisor's assets.

---

[185] *See* Vincent & Wilkins et al., *supra* note 182, at 56.

[186] *See Proving ownership of cryptocurrencies, Blockchain Innovation Group*, BLOCKCHAIN INNOVATION GRP., https://big-swiss.com/compliance-proving-ownership-of-cryptocurrencies/ (last visited Feb. 10, 2022) ("The most reliable way to prove ownership of crypto currencies is to sign a specified message with your [p]rivate [k]ey.") ("By doing so, the third-party can verify that the counterparty really knows the respective [p]rivate [k]ey without the need of revealing the very key or having to send a transaction.").

[187] Anchorage Letter, *supra* note 166; *see also supra* text accompanying notes 85, 93.

[188] *See* Vincent & Wilkins, *supra* note 182, at 54.

[189] *See id.*

[190] James Taylor, *Crypto.com Becomes the First Crypto Exchange to Comply with International SOC 2 Standards,* COINTRIBUTE (Nov. 24, 2021), https://www.cointribune.com/en/analysis/cybersecurity/crypto-com-becomes-the-first-crypto-exchange-to-comply-with-international-soc-2-standards/; *see also* Denis Omelchenko, *Gemini Custody Reports Successful SOC 1 Type 2 and SOC 2 Type 2 Audit,* IHODL (Jan 20, 2021), https://ihodl.com/topnews/2021-01-20/gemini-custody-reports-successful-soc-1-type-2-and-soc-2-type-2-audit/.

[191] *BitGo Leads Industry with Crypto Services SOC 2 Audit from Deloitte,* BLOOMBERG (July 20, 2018), https://www.bloomberg.com/press-releases/2018-07-20/bitgo-leads-industry-with-crypto-services-soc-2-audit-from-deloitte ("A SOC 2 report is designed to meet the needs of . . . clients who need assurances about the effectiveness of controls in place at a service organization . . . relevant to the . . . integrity of the system.").

Despite the unique nature of digital assets, blockchain technology presents a more efficient process for financial reporting and auditing.[192] As seen above, parties involved in the custody of digital assets are more than capable of meeting the requirements of auditing, verification, and segregation of digital assets.

c. Hard Forks, Airdrops, and 51% Attacks

A concern the SEC raised was the connection between hard forks, airdrops, and 51% attacks.[193] As a preliminary matter, it is important for the custodian and advisor to be accustomed with the underlying blockchain network their digital assets are supported on.[194] Since blockchain networks operate on different governance models[195] and mechanisms, the custodian should be able to assess and understand how network consensus is reached, how the blockchain is updated, and how the blockchain governance system functions.[196] The custodian should determine and assess potential vulnerabilities, such as hard forks, airdrops, and 51% attacks on the blockchains they intend to custody assets with.[197] The assessment of potential vulnerabilities should be ongoing, and the custodian should implement processes around the assessments of hard forks, 51% attacks, and airdrops.[198] It is imperative for the custodian to develop and enforce procedures and written policies in the event of blockchain malfunctions. While the mechanisms and approaches in which

[192] *Blockchain Technology and its Potential Impact on the Audit and Assurance Profession,* CHARTERED PRO ACCTS. OF CAN., https://us.aicpa.org/content/dam/aicpa/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/blockchain-technology-and-its-potential-impact-on-the-audit-and-assurance-profession.pdf (last visited Feb. 27, 2022).

[193] *Id.*

[194] Anchorage Letter, *supra* note 166.

[195] Jake Frankenfield, *On-Chain Governance*, INVESTOPEDIA, https://www.investopedia.com/terms/o/onchain-governance.asp (last updated Oct. 25, 2021) (noting on-chain governance refers to the system that the blockchain developers use for implementing and managing changes on the blockchain and includes the rules and procedures governing changes to the blockchain protocol).

[196] Anchorage Letter, *supra* note 166.

[197] *Id.* A custodian can determine how resilient the network is when facing a 51% attack through assessing the consensus mechanisms. *See id*.

[198] *Id.*

51% attacks and hard forks can be predicted is beyond the scope of this paper, there are currently commercial services available for market participants to actively monitor and manage public blockchain network risks.[199]

Before diving into the specifics of how custodians can handle blockchain challenges, it is important to note that hard forks, airdrops, and 51% attacks will not lead to the qualified custodian's assets being lost, stolen, nor do they adversely affect the private keys.[200] The events instead could cause disruption to the underlying blockchain network.[201]

i. Hard Forks

A hard fork occurs when one cryptocurrency splits in two, resulting in an old and new version of the blockchain.[202] In other words, after the fork, there are two parallel networks.[203] Up until the point where the fork occurred, the history of data is the same, but the two networks will

[199] CM Farum, COINMETRICS, https://coinmetrics.io/farum/ (last visited Feb 1, 2020))"Coin Metrics' Farum provides a comprehensive approach to blockchain network risk management."*)* (noting also the service allows user to identify 51% attacks, hard forks, and market abnormalities)*; see also 51% Attacks,* DIGITAL CURRENCY LAB: MIT MEDIA LAB, https://dci.mit.edu/51-attacks (last visited Feb. 1, 2022) ("We built a system to actively monitor scores of Proof-of-Work cryptocurrencies and detect chain reorganizations (reorgs) which can indicate that at 51% attack has occurred.").

[200] tZero Letter, *supra* note 159, at 16.

[201] *Id.*

[202] *Understanding Hard Forks in Cryptocurrency,* CRYPTOCURRENCY FACTS, https://cryptocurrencyfacts.com/understanding-hard-forks-cryptocurrency/ (last visited Feb. 1, 2022) (explaining since blockchains are not run by a centralized authority, the participants on the network (i.e., miners, node users, developers) vote on changes and upgrades to the network and forks make it possible for the networks to integrate new features as they are developed).

[203] *Hard Forks and Soft Forks Explained*, BINANCE ACAD., https://academy.binance.com/en/articles/hard-forks-and-soft-forks (last updated Feb. 18, 2022) [hereinafter BINANCE]; *see also Hard and Soft Forks: A Detailed and Simplified Explaination of How Blockchains Evolve,* FREEMAN L., https://freemanlaw.com/hard-and-soft-forks-a-detailed-and-simplified-explanation-of-how-blockchains-evolve/ (last visited Feb. 1, 2022) ("Because the two [networks do not] coexist, the new [network] splits into two branches, forming a fork-like diversion from the main [network].").

use different blocks and transaction afterward.[204] As a result of the shared history on the networks, the owner will have assets on both networks, assuming ownership before the hard fork occurred.[205] Let's say an owner has one Bitcoin when at block 600,000 at the time of the fork. The owner could redeem the Bitcoin on the old network in Block 600,001; in this case, the transaction would not occur on the new network's block 600,001.[206] In the case the owner does not sell or transfer the asset to the new network, the underlying protocol will continue to govern and this will not cause any changes to the nature of the digital asset.[207] Importantly, while two different networks are adopted and operate independently of one another, one network becomes dominant, which results in the other having low value and adoption.[208]

As mentioned earlier, custodians will have policies and procedures in place to monitor for hard forks and will have advance notice before the update is implemented. Upon notice of a hard fork, the custodian will notify its customer (i.e., the investment advisor) and inform the advisor of existing disclosures regarding this risk.[209] The advisor would need to determine which network to follow after the fork. The investment advisor can choose to keep the assets on the old network or transition to the new network.[210] During this time, the custodian will determine, based on its existing procedures, if it has the capability to custody the digital asset on the new network.[211] The

[204] *See* BINANCE, *supra* note 203.

[205] *Id.*

[206] *Id.*

[207] tZero Letter, *supra* note 159, at 16; *see also* Jamie Redman, *This Happens to Your Coins During a Bitcoin Hard Fork and Possible Blockchain Split,* BITCOIN.COM: News (Mar. 20, 2017), https://news.bitcoin.com/this-happens-to-your-coins-during-a-bitcoin-hard-fork-and-possible-blockchain-split/ ("The first and foremost piece of information all bitcoin holders should know is that in the event of a hard fork that splits the blockchain, bitcoins you possess will be perfectly safe.").

[208] *What are Forks and How do They Impact the Price of Cryptocurrency?,* COMMODITY.COM, https://commodity.com/cryptocurrency/what-are-forks/ (last updated Feb. 23, 2021).

[209] tZero Letter, *supra* note 159, at 14.

[210] *See How to Safely Claim any Cryptocurrency Fork,* CRYPTOCURRENCY FACTS, https://cryptocurrencyfacts.com/how-to-safely-claim-any-cryptocurrency-fork/ (last visited Feb. 8, 2022) (discussing how to move digital assets to the new network).

[211] tZero Letter, *supra* note 159, at 14.

custodian would also conduct an assessment on whether the new network has any operational or security problems and any material weaknesses.[212] In the event the custodian cannot custody the digital asset on the new network, policies and procedures would be in place to return the assets to the advisor, which would be communicated to the advisor in advance.[213]

Since hard forks do not affect custody of the digital asset as it can always remain on the old network, the only potential risk present is volatility in price.[214] In the event of a hard fork, there would not be added risks for an investment company vehicle directly holding digital assets compared to a current publicly available crypto-linked investment company vehicle. One such vehicle is BITO ETF, which uses futures contracts to track the price of Bitcoin.[215] Let's use BITO and a fictional ETF (ETF #2) that directly holds Bitcoin as an example. Both BITO and ETF #2 are operational when Bitcoin experiences a hard fork. BITO will continue to track the price of Bitcoin through its exposure to futures contracts and ETF #2 will continue to track the price of Bitcoin through direct custodianship. Investors will be exposed to the same amount of risk if they invest in BITO or ETF #2, assuming ETF #2 continues on the old network. As we can see, hard forks do not affect the custody of digital assets or private keys. Nor do hard forks introduce new risks to investment company vehicles more than those vehicles that have received the blessing of the SEC.

ii. Airdrops

[212] *Id.*

[213] *Id.*

[214] Redman, *supra* note 207.

[215] Ashworth, *supra* note 2.

An airdrop is a promotional activity where a digital asset is transferred to wallet addresses for free to help spread awareness of a particular digital asset.[216] Often airdrops are issued in exchange for completing specific tasks, usually to help increase the awareness of a particular crypto project.[217] Other times, the user of the wallet can choose to accept the tokens through claiming them.[218] The custodian can simply ignore such requests, therefore not receiving the digital asset.

In the case of an unexpected airdrop, the custodian would have policies and procedures to record unexpected airdrops.[219] While the custodian could choose to send back the asset to the sender or liquidate the asset, the custodian should avoid any interactions with the airdropped assets including approving or transferring the asset.[220] Phishing campaigns are known to take place through airdrops.[221] However, the custodian should not attempt to locate the sender or interact with the airdropped asset. Instead, the custodian or advisor should leave the airdropped digital assets in the wallet. As discussed in the auditing and verification section, the audit trail would track the added transaction.[222] The custodian would record this physically on the books, and the assets would be reconciled from the client's funds. The custodian and advisor would have written policies

[216] Jake Frankenfield, *Cryptocurrency Airdrop,* INVESTOPEDIA, https://www.investopedia.com/terms/a/airdrop-cryptocurrency.asp (last updated Feb. 13, 2022).

[217] Andrey Sergeenkov, *What is a Crypto Airdrop?,* COINDESK, (Jan. 18, 2022), https://www.coindesk.com/learn/what-is-a-crypto-airdrop/.

[218] Taylor Locke, *Crypto Projects are Increasingly Airdropping Free Tokens – But Investors Should be Cautions,* CNBC (Jan. 4, 2022, 11:21 AM) https://www.cnbc.com/2022/01/04/token-airdrops-are-common-in-crypto-but-investors-should-be-cautious.html.

[219] tZero Letter, *supra* note 159.

[220] *Security PSA: Airdrop Phishing Campaign,* COINBASE (Oct. 25, 2021), https://blog.coinbase.com/security-psa-airdrop-phishing-campaign-38b880c0298a [hereinafter *Security PSA*].

[221] *Id.*

[222] *See supra* text accompanying notes 179 – 184.

and procedures not to interact with such transactions. Because airdrops can only cause harm through interaction with the airdropped assets, they would not pose a risk to the client's assets. [223]

iii. 51% Attacks

A 51% attack occurs when malicious users control more than 50% of a blockchain's mining capabilities (i.e., hashing power).[224] The more hashing power a digital asset's underlying network generates, the greater security against a 51% attack.[225] Less established digital assets are more susceptible to 51% attacks due to their low hashing power.[226] This is not the case for larger established digital assets, such as Bitcoin or Ethereum.[227] "Theoretically, 51% attacks could, but are unlikely to[] impact large cryptocurrencies because the possibility of an individual or group controlling greater than half of all mining or validation power of . . . Bitcoin is so unlikely as to be essentially impossible."[228] As with other potential concerns regarding vulnerabilities to blockchain networks, advisors for investment companies and custodians would review the digital asset's network concerning 51% attacks before deciding to invest in or custody the asset.[229] Digital assets such as Bitcoin or Ethereum would be appropriate for the reasons explained above. Digital assets

[223] *See Security PSA*, *supra* note 220; *see also How to Avoid Airdrop Scams,* MEDIUM (June 6, 2018), https://medium.com/@AirdropReview/how-to-avoid-airdrop-scams-ce3eaaa94add.

[224] Ephraim Njoroge, *Understanding a 51% Attack on the Blockchain*, SECTION (Dec. 15, 2021), https://www.section.io/engineering-education/understanding-the-51-attack-on-blockchain/.

[225] *See* Jimi S., *Blockchain Explained: How a 51% Attack Works*, GOOD AUDIENCE (May 5, 2018), https://blog.goodaudience.com/what-is-a-51-attack-or-double-spend-attack-aa108db63474.

[226] *Id.* "[Lack of hashing power] is why 51% attacks usually occur on small blockchains . . . if they occur at all." *Id.*

[227] *See* Njoroge, *supra* note 224.

[228] Eric Rosenberg, *What is a 51% Attack*, THE BALANCE (Oct. 28, 2021), https://www.thebalance.com/what-is-a-51-attack-5207550#:~:text=Theoretically%2C%2051%25%20attacks%20could%2C,as%20to%20be%20essentially%20impossible; *see also How Much Would it Cost to 51% Attack Bitcoin?,* BRAIINS: BITCOIN MINING INSIGHTS (Jan. 11, 2021), https://braiins.com/blog/how-much-would-it-cost-to-51-attack-bitcoin ("There have been no successful 51% attacks on Bitcoin."); *see also* tZero Letter, *supra* note 159, at 19 ("Expected network updates to Ethereum . . . will further reduce the probability of a 51% attack because it is transitioning to a proof-of-stake network (in such a case, one would have to acquire over 51% of a network's staked non-security digital assets to successfully launch a 51% attack.").

[229] *See* tZero Letter *supra* note 159, at 19.

supported on networks that are vulnerable to 51% attacks would not be appropriate for advisors to invest in or for custodians to custody.[230]

In the event a 51% attack is successful, the attackers are not able to create new coins on the network, but they are able to "double spend"[231] coins. Attackers can also participate in "selfish mining," which prevents other miners from carrying out their function where the attackers can generate empty blocks which could stall the network.[232] The disruption a 51% percent attack can cause is limited.[233] The attackers cannot prevent other users from broadcasting their transactions on the network or reverse transactions.[234] Attackers can also change the order in which transactions occur.[235] This type of attack is incapable of creating new assets or stealing the assets of other users on the network.[236] With this said, the result of a 51% attack does not change the nature of the digital assets, the custodian's or advisor's custody of the assets nor do these attacks compromise the private keys.[237]

Because investment companies, advisors, and custodians would only invest in and custody assets with established crypto projects supported on networks with strong protection against 51%

[230] *See How to Prevent a 51% Attack,* 2MINERS.COM (Apr. 3, 2019), https://2miners.com/blog/how-to-prevent-a-51-attack/ (noting that the choice of digital assets for advisors and custodians may not be limited to established ones, i.e., new technologies such as Komodo's Delayed Proof of Work is a universal solution that fits all digital asset networks with low hash rates and fights against 51% attacks).

[231] Matthew Warholak, *Bitcoin's Double Spending Problem: Definition, Solution, and Future Outlook,* SOFI (Mar. 15, 2021), (explaining a double spend occurs when a bad actor sends a copy of one transaction making the copy appear legitimate while retaining the original transaction resulting in a single digital currency being spent simultaneously more than once).

[232] *The History of 51% Attacks and the Implication for Bitcoin,* MEDIUM (May, 7, 2019), https://medium.com/hackernoon/the-history-of-51-attacks-and-the-implications-for-bitcoin-ec1aa0f20b94.

[233] Griffin McShane, *What is a 51% Attack,* COINDESK (Oct. 12, 2021), https://www.coindesk.com/learn/what-is-a-51-attack/#:~:text=The%2051%25%20Attacks!&text=Successful%20attackers%20gain%20the%20ability,issue%20known%20as%20double%20spending.

[234] *Id.*

[235] *What Everyone Gets Wrong About 51% Attacks,* DANKRAD FEIST (May 20, 2021), https://dankradfeist.de/ethereum/2021/05/20/what-everyone-gets-wrong-about-51percent-attacks.html.

[236] McShane, *supra* note 233.

[237] tZero Letter, *supra* note 159, at 19 – 20.

attacks, the SEC's concern of 51% attacks is alleviated. Moreover, because 51% attacks do not affect the private keys and ultimately, the custody of digital assets, illustrates that these assets are safe for custody.

CONCLUSION

While digital assets are a relatively new, but evolving asset class, market participants are adopting and pushing forward innovation in the space. Congress and the SEC certainly could not have foreseen the invention of an asset that is completely intangible when enacting the ICA and the IAA or their accompanying rules. Despite this, the safeguards Congress and the SEC established to protect investors can be fulfilled regarding investment companies and associated parties directly holding digital assets. Today, there are clear, safe, and proven mechanisms that allow for investment vehicles such as ETFs to directly hold digital assets. Custody has been the missing piece of digital asset infrastructure. Resolution of the custody issue is the next step towards digital assets being seen as a safe financial asset and for market confidence more generally. Until regulators resolve the issue of custody, the inability for participants in this sphere to directly hold digital assets will deprive society of the benefits digital assets have to offer.