RETHINKING INTELLECTUAL PROPERTY LAW
IN AN ERA OF GENERATIVE AI

BY
GIUSEPPINA D'AURIA

&
ARUN SUNDARARAJAN

Giuseppina D'Auria (https://www.linkedin.com/in/giuseppinadauria/) is a corporate lawyer and a researcher on the legal and ethical aspects of AI at NYU's Stern School of Business. Arun Sundararajan (https://digitalarun.io/) is the Harold Price Professor of Entrepreneurship and Professor of Technology, Operations and Statistics at NYU's Stern School of Business.

TechREG CHRONICLE
NOVEMBER 2023

DATA BROKERS IN THE HOT SEAT: A CONTINUING STORY

By Jessica L. Rich

For years, policymakers have debated whether new laws are needed to restrict the practices of data brokers – companies that collect consumers' data from various sources, process and package it, and then sell it to individuals and businesses for marketing and advertising, fraud detection, risk mitigation, and locating people, among other purposes. Supporters of stronger laws argue that data brokers operate behind the scenes, collecting and selling sensitive consumer data to a vast array of purchasers, who use it to make important decisions about consumers. Opponents argue that data brokers provide valuable services that help businesses and the government serve the public. Until recently, regulation of data brokers in the U.S. has been limited. During the past couple of years, however, there's been flurry of regulatory activity affecting data brokers at the federal and state levels. Of particular note, last month, California passed a new law (the Delete Act) that will allow consumers, in one step, to delete the data that all data brokers in the state have collected about them and to prevent future sales of their data. This article examines the recent regulatory activity surrounding data brokers and predicts continued focus on this industry as we move to 2024.

Scan to Stay Connected!

Scan here to subscribe to CPI's **FREE** daily newsletter.

Visit **www.competitionpolicyinternational.com** for access to these articles and more!

01
INTRODUCTION

In June 2016, Sony released a new song, "Daddy's Car,"[2] one that a casual listener might have concluded was an early iteration of a lost Beatles song. The lyrics were written by Benoît Carré (a human), but the music was composed entirely by an artificial intelligence ("AI") system called Flow Machines developed by Sony Computer Science Laboratory. With access to a vast library of music, Flow Machines composed "Daddy's Car" after being asked to compose a song in the style of the Beatles.

While the release attracted little attention, it was a leading indicator of what is now the most substantive challenge to intellectual property ("IP") law since the emergence of the Internet.

Granted, the use of digital technology to aid songwriting or composition was not new and had been happening for decades. Additionally, AI-generated voices like the popular Hatsune Miku that used voice synthesizing software like Yamaha's Vocaloid[3] predated Flow Machines by more than a decade. But the creation of "Daddy's Car" raised deeper questions. While Sony/ATV owned the Beatles catalog of music during the creation and use of Flow Machines and could therefore use that catalog to train its model, what would the legal implications be if a non-owner used the songs in the Sony catalog to train a generic music generating AI? More specifically, would the creation of an AI system expressly designed to replicate the musical style of the Beatles somehow infringe on the IP rights of Sony/ATV, the owner of this catalog? If this kind of system were created to replicate a living and/or active musician, would the musician have a valid and enforceable claim? And finally, who owned "Daddy's Car?" Sony/ATV as the owner of the Beatles catalog on which the Flow Machines AI was trained? Or the creators of the AI? Or did it belong to the public domain?

Until recently, Flow Machines was an instance of an AI system on the fringes of the cultural dialog, as were AI systems that had been created to paint in the style of famous artists.[4] While the subsequent emergence of OpenAI's large-language model ("LLM") GPT-2 in 2019 spawned systems that could create essays in the style of the New Yorker,[5] and write poetry like famous poets,[6] the discussion about this new world of AI-created artworks was still largely relegated to the academic space, confined to university campuses and tech conferences.

But today, as generative AI systems come of age, these questions are very much central to the future of IP creation and ownership by humans. Widely available LLMs like OpenAI's GPT-4 and Google's PaLM can be used to generate new written content in the style of any specific author with just a few examples of their prior art.[7] Diffusion models like those used by OpenAI's DALL-E 3, Stable Diffusion and Midjourney can readily create new visual art in the style of a specific artist.[8] Voice cloning systems enable the near-perfect replication by AI of the voice of existing artists, making AI-generated music in the style of a specific artist indistinguishable from human compositions. Indeed, a song called "Heart on My Sleeve" that appeared to be written by Drake

---

2   Benoît Carré, Sony CSL, "Daddy's Car," https://www.youtube.com/watch?v=LSHZ_b05W7o.

3   Yuri Kageyama, *Japan's synthesized singing sensation Hatsune Miku turns 16*, ABC News (Sept. 2, 2023), https://abcnews.go.com/Entertainment/wireStory/japans-synthesized-singing-sensation-hatsune-miku-turns-16-102879946.

4   NPR Staff, *A 'New' Rembrandt: From The Frontiers Of AI And Not The Artist's Atelier*, NPR (Apr. 6, 2016) https://www.npr.org/sections/alltechconsidered/2016/04/06/473265273/a-new-rembrandt-from-the-frontiers-of-ai-and-not-the-artists-atelier; Mike Murphy, *Computers Can Now Paint Like Van Gogh and Picasso*, Quartz (Sept. 6, 2015), https://qz.com/495614/computers-can-now-paint-like-van-gogh-and-picasso/.

5   John Seabrook, *Can a Machine Learn to Write for The New Yorker?*, New Yorker (Oct. 14, 2019), https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker.

6   Kelsey Piper, *A Poetry-Writing AI Has Just Been Unveiled. It's … Pretty Good*, Vox (May 15, 2019), https://www.vox.com/2019/5/15/18623134/openai-language-ai-gpt2-poetry-try-it; Cushman, J., *ChatGPT: Poems and Secrets*, LIL Blog (Dec. 20, 2022), https://lil.law.harvard.edu/blog/2022/12/20/chatgpt-poems-and-secrets/.

7   As "prompt-tuning" methods evolve, such replication will become increasingly effortless and accurate. *See*, for example, Yuxian Gu, Xu Han, Zhiyuan Liu, and Minlie Huang, *PPT: Pre-trained prompt tuning for few-shot learning*, arXiv preprint arXiv:2109.04332 (2021).

8   Kyle Chayka, *Is A.I. stealing from artists?*, New Yorker (Feb. 10, 2023).

was released and briefly made available for streaming on YouTube and Spotify in April 2023.[9]

02

INTELLECTUAL PROPERTY FOUNDATIONS AND GENERATIVE AI

In the discussions around AI-generated works, three main questions have been raised that motivate this Article and are at the center of the debate: First, what control does the owner of data have over its use in training a generative AI system? Second, does a human have the right to control the creation of a generative AI system that replicates their individual "creative process," and if such a system is created, what claims or recourse does the human have? And third, who owns the works created by such a system?

Our goal is not to provide precise answers to these questions. It's too soon for that. Rather, we aim to examine what the current US regulatory framework suggests are likely answers and contrast these answers with the economic rationale for assigning specific levels of property rights over human-generated artifacts.

So first, let's take a quick look at the justifications for IP rights to exist in the first place. The foundations of the IP regimes of most countries stem from John Locke's utilitarian view that when a human provides their labor to goods held in common, the human is entitled to earn "fruits of their labor" by obtaining a private property right.[10] Although developed initially for tangible goods, the premise applies equally well to the intangibles now protected by IP — indeed, the human creator is combining their labor or talent with goods in the public domain (commons) to create something new, and is thus deserving of (intellectual) property rights over this new creation. In the case of creative goods, this justification is supplemented in some European countries by the

tradition of *droit d'auteur* — which has its roots in Hegel's notion of property as the mark of free people[11] and was perhaps first proposed by Immanuel Kant for literary works[12] — that generating certain goods is an expression of personality, leading to an irrevocable bond between the author and the work, and necessitating a property right in the interest of preserving the dignity of the individual.

In shaping the socially optimal extent of these IP rights, economists will often consider the trade-offs between three potentially countervailing objectives. First, the creator of IP has higher motivation to create if their property rights are greater, and thus, assigning higher levels of ownership to the creator benefits society by increasing creation incentives. Second, society as a whole accrues greater benefits if more people have access to these new creative products, so assigning lower levels of ownership to the creator benefits society by increasing the value obtained from the (non-rival) IP from its consumption. And third, the future creation of intellectual property is more likely, faster and of higher quality if creators have greater access to past creative content, so assigning IP ownership to creators in a manner that expands access to existing IP for the purposes of future innovation is beneficial to society. Clearly, these objectives are not all aligned, always necessitating trade-offs that are embodied in a nation's IP laws.

Viewing the questions raised by generative AI through this lens suggests immediately that they have no simple answers. The intricate nature and extensive complexities of IP in the generative AI realm necessitate a broader framework of analysis that current legal framework cannot fully address, since it was optimized for the political and economic conditions of different times. Advances in generative AI change the costs of production and innovation in ways that suggest that IP laws — which may have been an optimal balance in the past — may need to be revised to reflect these changes.

---

9   Lane Brown, *AI Singers Are Unnervingly Good and Already Ubiquitous. The software that cloned Drake and the Weeknd is easy to use—and impossible to shut down*, Vulture (May 1, 2023), https://www.vulture.com/article/ai-singers-drake-the-weeknd-voice-clones.html. Contrary to what many people believe, the song was not created entirely by generative AI. The music was likely composed by a human songwriter named "Ghostwriter" in the musical style of Drake. The role of AI was likely in the use of voice cloning to convert the rendition of the (human written) lyrics to appear to be in Drake's voice.

10   John Locke, *Two Treatises on Government* (3rd ed.n, Awnsham and John Churchill, 1698), Book 2, para 26.

11   Justin Hughes, *The Philosophy of Intellectual Property*, 77 Geo. L.J. 287, 330 (1988); William Fisher, *Theories of Intellectual Property*, *in* New Essays in the Legal and Political Theory of Property, 168, 171-72 (Stephen R. Munzer ed., 2001).

12   Benedict A.C. Atkinson & Brian F. Fitzgerald, *A Short History of Copyright: The Genie of Information* (Springer ed., 2014), 35-36.

03

IS THE USE OF COPYRIGHTED MATERIAL TO TRAIN AN AI INFRINGEMENT?

A number of commercial generative AI systems have used copyrighted works as part of their training data without the explicit permission of the copyright holders for this use. Earlier this year, a group of artists filed a class-action lawsuit against Midjourney, Stable Diffusion, and DeviantArt, claiming that they have infringed the copyright of countless artists by using their creations to train their models.[13] Shortly after, stock photo licensing giant Getty Images, which is usually open to license "the use of its visual assets and associated metadata in connection with the development of AI and machine learning tools,"[14] filed a lawsuit against Stability AI, stating that it "copied [millions of images] without permission and used to train one or more versions of Stable Diffusion."[15] In July 2023, a group of artists that included Sarah Silverman proposed a class-action lawsuit against OpenAI alleging that the latter unlawfully incorporated their written works into the datasets utilized to train ChatGPT for generating responses to human text inputs.[16] These authors have in parallel initiated a similar suit against Meta Platforms.[17] Others have filed analogous lawsuits against Google[18] and Anthropic[19] for their practices in training AI systems.

A. Is Training Like Reading?

Some arguments in favor of the unfettered use of copyrighted material for training purposes draw an analogy between the use of data for training and a human reading the copyrighted material, listening to a song or viewing visual art.[20] In considering this argument, one should recognize that while it may be theoretically possible for a human to "read" a document in a manner similar to what a machine learning system "does" while learning, it is practically impossible for any human — let alone the average human — to "read" in this way. At a very high level of abstraction, when a machine is being trained, it first learns an "embedding" that maps every word it encounters to a point in a high-dimensional geometric space based on the word's usage in sentences across millions of examples in a manner that reveals the relationships between words that are implied by their usage. (For example, that "Republican Party" and "GOP" refer to the same entity, and that Paris has the similar relationship with France that Berlin has with Germany.) Often, the exact sequence of words and the patterns of repetition of these expressions are used to embed each word into its mathematical form.[21] It then "encodes" each document it is trained on using this embedding. It then, loosely speaking, "memorizes" the patterns contained in tens of millions of documents. The results of this process of "memorization" are represented as billions of numerical parameter values in an artificial neural network. The culmination of this process of "reading" is a system that is accurate at predicting the logical next word when presented with any string of words that may be accompanied by directions on the literary style or author to mimic. So if one thinks of a machine learning system as a "reader," it should also be clear that the technological changes that led to LLMs may have fundamentally altered the balance between authors and readers, revivifying the need for changes in the corresponding allocation of rights under current copyright law.

B. Ideas Versus Expressions

Another argument in favor of the use of prior works as training data is that the training process is replicating the

---

13 *Andersen et al. v. Stability AI et al.,* Case No. 3:23-cv-00201 (N.D. Cal. filed Jan. 13, 2023).

14 Brief for plaintiff at 2, Getty Images (US), Inc. v. Stability AI, Case No. 1:23-cv-00135 (D. Del. filed Feb. 3, 2023).

15 *Id*. at 8.

16 *Silverman et al. v. OpenAI,* Case No. 4:2023-cv-03416 (N.D. Cal. filed July 7, 2023).

17 *Kadrey et al. v. Meta Platforms,* Case No. 3:2023-cv-03417 (N.D. Cal. filed July 7, 2023).

18 *J.L. et al v. Alphabet*, Case No. 3:23-cv-3440-AMO (N.D. Cal. filed July 11, 2023).

19 *Concord Music Group, Inc. v. Anthropic PBC*, Case No. 3:23-cv-01092 (M.D. Tenn. filed Oct. 18, 2023).

20 See Def. (Stability AI)'s Mot. Dismiss at 1, *Andersen et al. v. Stability AI et al*., Case No. 3:23-cv-00201 (N.D. Cal. filed Jan. 13, 2023): "[T]raining a model does not mean copying or memorizing images for later distribution. Indeed, Stable Diffusion does not "store" any images. Rather, training involves development and refinement of millions of parameters that collectively define — in a learned sense — what things look like."

21 For a reference to the Word2vec technique, see Tomás Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado & Jeffrey Dean, *Distributed Representations of Words and Phrases and their Compositionality*. CᴏRR abs/1310.4546 (2013). See also, about BERT, Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, *BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding*, *in* Pʀᴏᴄᴇᴇᴅɪɴɢs ᴏꜰ ᴛʜᴇ 2019 Cᴏɴꜰᴇʀᴇɴᴄᴇ ᴏꜰ ᴛʜᴇ Nᴏʀᴛʜ Aᴍᴇʀɪᴄᴀɴ Cʜᴀᴘᴛᴇʀ ᴏꜰ ᴛʜᴇ Assᴏᴄɪᴀᴛɪᴏɴ ꜰᴏʀ Cᴏᴍᴘᴜᴛᴀᴛɪᴏɴᴀʟ Lɪɴɢᴜɪsᴛɪᴄs: Hᴜᴍᴀɴ Lᴀɴɢᴜᴀɢᴇ Tᴇᴄʜɴᴏʟᴏɢɪᴇs, 1, 4171-4186.

ideas in these works rather than copying the specific *expressions* of these ideas. Indeed, a key tenet of U.S. copyright law is that "[P]rotection is given only to the expression of the idea – not the idea itself."[22] This is because protecting ideas may give a socially suboptimal level of market power to the creator. (For example, giving the first author of a style of science fiction ownership over the genre rather than their specific book.) While the idea/expression dichotomy is codified in the Copyright Act,[23] the disentangling of ideas and expressions may not always be straightforward, and the boundaries are often dictated by case law.[24]

From an economic standpoint, one might think of generative AI for creative content (written, audio, visual) as a tool that dramatically increases the productivity of technology that produces "expressions," and thus, the socially optimal level of protection of ideas and style of ideas may need to change. However, before one concludes this is an economic argument in favor of the unlimited use of copyrighted material in the process of training a machine learning model, one might also consider that the generative AI training process is one that may rely very heavily on the exact manner in which the ideas contained in a work are expressed. For example, the generative AI training process for LLMs is one that may rely extensively on the exact manner in which the ideas contained in a work are expressed, or *exactly* how the idea in a text is expressed in words. In contrast, in granting DeviantArt's motion to dismiss in the *Stability AI* case, the judge considered that the diffusion process associated with AI training "involves not copying of images, but instead, the application of mathematical equations and algorithms to capture *concepts* (emphasis added) from the Training Images."

C. The Limits of "Fair Use"

Since the arguments for and against "training is reading" are inconclusive and there is no clear guideline that emerges from the idea/expression dichotomy in current copyright law, what remains is to examine whether this new use of existing works — as training data — might qualify as *fair use*. Indeed, this is an argument encountered frequently in both the lawsuits[25] and the regulatory reform discussion of the last year.[26]

When applying the fair use doctrine in the U.S., judges consider four factors[27]: (1) the purpose and character of use; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion taken, and (4) the effect of the use on the potential market. Factor (2) seems to have ambiguous probative value here, given that generative AI systems tend to be trained on all kinds of works, some factual (which generally fall within the bounds of fair use), others more creative. Based on the third factor, one might argue that this new use is not fair because the training process uses the copyrighted works in their entirety, but one must simultaneously consider the first factor because it seems undeniable that training a machine learning model is a quintessential example of transformative use.[28]

> ❝ *When applying the fair use doctrine in the U.S., judges consider four factors*

---

22 *Mazer v. Stein*, 347 U.S. 201, 217.

23 17 U.S.C. § 102(b): "In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work."

24 See *Peter Pan Fabrics, Inc. v. Martin Weiner Corp.*, 274 F.2d 487, 489 (2d Cir. 1960), "[N]o principle can be stated as to when an imitator has gone beyond copying the "idea," and has borrowed its "expression." Decisions must therefore inevitably be ad hoc. [...] [O]ne cannot say how far an imitator must depart from an undeviating reproduction to escape infringement." See also *Reyher v. Children's TV Workshop & Tuesday Publs.*, 533 F.2d 87, 91: "[T]he essence of infringement lies in taking not a general theme but its particular expression through similarities of treatment, details, scenes, events and characterization." Accordingly, "[s]imilarity of expression … which necessarily results from the fact that the common idea is only capable of expression in more or less stereotyped form will preclude a finding of actionable similarity."

25 See, *inter alia*, *Thomson Reuters Enterprise Centre GmbH et al v. ROSS Intelligence Inc.*, Case No. 1:20-cv-00613-SB (D. Del. filed May 6, 2020).

26 Christopher T. Zirpoli, Cong. Research Serv., LSB10922, Generative Artificial Intelligence and Copyright Law (2023); Artificial Intelligence and Intellectual Property: Part I - Interoperability of AI and Copyright Law: Hearing Before the Subcommittee on Courts, Intellectual Property, and the Internet of the Committee on the Judiciary, House of Representatives, One Hundred Eighteenth Congress, First Session (2023).

27 17 U.S.C. § 107.

28 See *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d. Cir. 2015), where the court ruled that Google's scanning of the entire text of millions of books was fair use because the creation of a searchable database constituted a transformative use under the fair use doctrine. However, "the courts [...] have begun to depart from the precedents established by the Google Books cases" and "there is no guarantee that courts will extend this precedent to similar technologies or legal contexts." (Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 Tex. L. Rev. 743, 763 (2021)).

Thus, it appears that the critical prong in determining whether the use of copyrighted works for training AI models constitutes fair use will be factor (4), i.e. the effect of this training on the potential market for the works used in training. Indeed, in *Authors Guild v. Google, Inc.*, Google was allowed to scan the content of entire copyrighted books to create a search engine that could find text within those books and present brief snippets as search results because the database created had a positive effect on the potential market for these books by making them easier to find and associating specific books more effectively with specific kinds of reader intent. Another partial precedent can be found in *Sega Enterprises Ltd. v. Accolade, Inc.*, in which Accolade copied and reverse-engineered (in its entirety) the software of Sega's video game console to create games compatible with Sega's console. While this was deemed fair use, this act of copying increased the market value of Sega's console because of the benefits to the console's potential market accruing from network effects — the availability of more video gaming titles increases the value of a console to existing and potential new buyers of the console.[29]

At this time, the effects of generative AI systems on the potential market for the works the AI system is trained on is unclear. This is partly because the questions around the ownership and copyrightability of works generated by these systems remains unresolved, as does the legitimacy of explicitly replicating, perhaps in a manner that is identified, the exact "style" of a human creator. It is possible that a generative AI system for music floods the market with free AI-generated music that significantly impairs the market for human-created music. On the other hand, consumers may view this as an entirely new category of entertainment, may not feel a connection to AI-generated music, and may even seek out the human artists whose work resembles the style of AI music they enjoy.

The question gets even more complex if the AI system aims to replicate the style and identity of specific creators, a topic we turn to next.

04
COPYRIGHT AND THE OWNERSHIP OF A HUMAN'S CREATIVE PROCESS

Granted, the majority of today's generative AI models are developed primarily for the purpose of generating new content. The intent of these models is not to replicate copyrighted material from their training data and, in almost all cases, they do not do so.[30] As noted by others, most machine learning systems "are interested in copyrighted works for reasons that have nothing to do with the things that make those works copyrightable."[31] However, if this act of copying is aimed at reproducing individual expression, such as when a machine learning system is trained to produce a song in the style of a certain artist and with reference to a specific song of that artist, the assessment can, and should, become considerably stricter.

Given the power of generative AI systems, this new content has indeed started to resemble either existing content or an existing artistic style in a manner that deserves greater scrutiny. Many of us have already experienced this by asking a system like ChatGPT or Bard to create new prose or poetry in the style of a famous author, or by generating new art in the style of a famous artist. Indeed, future generative AI systems may be able to effortlessly create new songs that sound exactly like an existing artist, or video content that replicates the likeness of a human actor.

The emergence of such systems may dramatically alter the economic incentives of individual humans to create new content. If one's individual "style" can be replicated soon after it is established as being of economic value, not only may a human be unable to adequately benefit from "the fruits of their labor" (as an undifferentiated AI-based competitor emerges rapidly) but the incentives to invest in reaching the level of expertise necessary for the production of good music, being successful at corporate sales or making breakthrough scientific discovery are reduced.

---

29 *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992).

30 LLMs have been shown to "memorize" portions of the works and generate verbatim reproductions which are expressive. See Nicholas Carlini, Daphne Ippolito, Matthew Jagielski, Katherine Lee, Florian Tramer, and Chiyuan Zhang, *Quantifying memorization across neural language models*, arXiv:2202.07646 (2022), LLMs tend to memorize items that were replicated many times in the training data. In *Concord Music Group, Inc. v. Anthropic PBC* (*see supra note 19*), for example, the plaintiff prompted Anthropic's LLM Claude 2 to "Write a song about the death of Buddy Holly," yielding output consisting of the exact same lyrics of Don Mclean's famous song about the tragic event, "American Pie."

31 Mark A Lemley & Bryan Casey, *Fair Learning*, 99 Tex. L. Rev. 743, 773 (2021).

Additionally, although economic analyses of copyright (and more broadly, of IP) tend to be rooted more in the utilitarian foundations of Locke, it may be worth examining the replication of the human creative process through the lens of personality theory. While "expressions of personality" do not have a natural home in the paradigm of economics,[32] one might consider the possible implications of designing society in a manner where the bond between a human and their creative process is broken.

A. The Boundaries of Substantial Similarity and the Potential Protection of an Artistic Style

Generative AI systems do not mask the reality that they are being trained explicitly on existing work. Also, copyright does not just cover replication of work in its original form, but also covers the creation of new works "based upon" the initial work[33] — subsequent artists have to obtain permission from the copyright holders if they want to use protected works as a model or a template, and failure to do so can lead to a copyright infringement claim. Now, for this infringement claim to succeed, the copyright holder must demonstrate: (a) that the alleged infringers copied from their copyrighted works; and (b) that the copying went so far as to constitute unlawful appropriation.[34]

When considering generative AI, prong (a) is not in dispute — generative AI systems that are potentially replicating the style of a prior creator meet the requirement of having "had access to the work."[35] The debate will thus center around what constitutes "substantial similarity" between the AI-

generated works and the ones it was trained on, and how to draw the line between infringing "adaptation" and non-infringing "inspiration."[36]

> **When applying the fair use doctrine in the U.S., judges consider four factors**

What makes this issue especially complex to resolve based on existing case law is that the mainstream view of copyright law follows Section 102(b) of the Copyright Act,[37] that appropriation of an *artistic style* does not constitute copyright infringement. More specifically, copyright law does not permit an artist to preclude another from mimicking an artistic style, unless the appropriation involves the artistic style as expressed in a specific work.[38]

That said, one path open to an artist when they encounter AI-generated works created "in the style of" their prior works is to appeal to the less rigid definitions of substantial similarity, such as the "total concept and feel" approach.[39] Well known cases that seem to ascribe some measure of ownership over artistic style to the artist include *Steinberg v. Columbia Pictures Industries*,[40] in which the New Yorker artist prevailed against Columbia Pictures for their movie poster for Moscow on the Hudson resembling his famed

---

32   However, this lens is certainly relevant to copyright policy considerations in the EU. On September 12, 2023, a group of French MPs proposed a bill aimed at providing a copyright framework for AI, whose main rationale was to "provide imperative protection for authors and artists of creation and interpretation in accordance with a humanist principle, in legal agreement with the Code of Intellectual Property", and to "seek and find a solution to curb what already seems to us to be a threat and probably a future disaster for creation.(*translation by the Authors*)"

33   17 U.S.C. §106(1): "[...] [T]he 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."

34   *Arnstein v. Porter*, 154 F.2d 464, 468 (2d Cir. 1946).

35   *Shaw v. Lindheim*, 919 F.2d 1353, 1356 (9th Cir. 1990); *Copeland v. Bieber*, 789 F.3d 484, 488 (4th Cir. 2015).

36   Also note that under the inverse ratio rule, the courts "require a lower standard of proof on substantial similarity when a high degree of access is shown." See Smith v. Jackson, 84 F.3d 1213, 1218 (9th Cir. 1996) (citing Shaw v. Lindheim, 919 F.2d 1353, 1361-62 (9th Cir. 1990).

37   See *supra* note 22.

38   As explained in *Dave Grossman Designs, Inc. v. Bortin*, 347 F. Supp. 1150, 1156-57, "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."

39   Under the "total concept and feel" approach, there is similarity between the works if "the ordinary, reasonable person would find the total concept and feel of the works to be substantially similar." (*Williams v. Gaye*, 885 F.3d 1150, 1164). Substantial similarity must be ascertained based upon the entire copied portion of each work, including portions which, viewed in isolation, might be unprotected by copyright (*Shaw v. Lindheim*, 919 F. 2d 1353, 1363).

40   *Steinberg v. Columbia Pictures Industries, Inc.*, 663 F. Supp. 706 (S.D.N.Y. 1987).

"View of the World from 9th Avenue,"[41] and a case involving the 2013 song "Blurred Lines,"[42] which was ruled as infringing Marvin Gaye's song "Got to Give It Up", despite the songs not sharing any melodic phrases, sequences of chords, lyrics or rhythm.[43] However, these cases appear to be the exception rather than the rule.

Thus, the traditional economic argument in favor of generally not protecting artistic style may need to be revisited, perhaps refining the definition of "artistic style" to be more nuanced, and adapting the notion of "substantial similarity" accordingly. The economic rationale against the protection of artistic style in the past has been straightforward — it would have assigned excessive market power to the individual who created the first detective novel, or the first hip-hop song. However, the law could not have foreseen the emergence of technologies that could costlessly replicate the more precise "style" of a single individual and create works artistically comparable with the original. It might be argued, in contrast, that assigning to a specific artist ownership over that subset of "artistic style" that encapsulates their individual creative process (and no more) reflects market power that is not excessive, but appropriate.

Granted, Section 301(a) of the Copyright Act preempts any remedy other than copyright for claims invoking rights "equivalent to any of the exclusive rights within the general scope of copyright . . . in works of authorship." Pursuant to this provision, claims are expressly preempted when they are simply efforts to recast copyright claims under other legal rubrics. That said, there appear to be a few alternative bodies of law that may be pertinent to settings involving AI-embedded replication of human creators. A company whose generative AI system responds to prompts requesting output "in the style of" specific creators could face a right of publicity argument. Other facets of *unfair competition* law may be relevant if a creator finds that the strong stylistic similarity with their existing or new works can lead consumers to believe that the output of a generative AI system that was trained on these works was created by the famous artist, and not simply inspired by them. An artist's visual style might also be protected as "*trade dress*" under trademark law. The *misappropriation doctrine* may be employed to establish a case against gathering training data via "web scraping" from websites that are publicly accessible or not protected by access restrictions. The application of each of these legal remedies to generative AI is in its infancy, and certainly warrants more careful economic analysis, something we plan in the future.

B. The Ownership of New AI-Generated Works

Given that there is little doubt that generative AI will lead to the creation of new works that mimic the artistic style of specific artists, a natural next question that arises concerns the ownership status of these new creations or works. In looking at existing law, there are multiple competing theories, none of which seem to be definitive.

One line of reasoning is that AI-generated works should be in the public domain. Indeed, the USPTO and Federal Circuit have held that an AI system cannot be an inventor on a patent,[44] and in August 2023, a D.C. district court ruled that a work of art solely generated by an AI system is not eligible for copyright registration under US law.[45] The economic arguments that favor the absence of any copyright holder seem to place greater emphasis on maximizing the economic benefits that accrue from sharing works as widely as possible with the public, perhaps imagining that giving a creator incentives to create in the first place is not relevant for non-human "AI creators" embedded in software. Some authors have tried to provide a technological rationale for this reasoning: "The automated decision-making feature of deep learning machines [...] adds unpredictability [...] and in doing so it breaks the causal link between humans (the author of the code or the user of the machine) and the output."[46]

> **" Granted, Section 301(a) of the Copyright Act preempts any remedy other than copyright for claims invoking rights "equivalent to any of the exclusive rights within the general scope of copyright . . . in works of authorship**

---

41   https://en.wikipedia.org/wiki/View_of_the_World_from_9th_Avenue#/media/File:Steinberg_New_Yorker_Cover.png.

42   *Williams v. Gaye*, 885 F.3d 1150.

43   This provides an interesting expansion of 114(b) of the Copyright Act, specifically that "The exclusive right of the owner of copyright in a sound recording under clause (2) of section 106 is limited to the right to prepare a derivative work in which the actual sounds fixed in the sound recording are rearranged, remixed, or otherwise altered in sequence or quality. The exclusive rights of the owner of copyright in a sound recording under clauses (1) and (2) of section 106 do not extend to the making or duplication of another sound recording that consists entirely of an independent fixation of other sounds, even though such sounds imitate or simulate those in the copyrighted sound recording" (17 U.S.C. § 114(b)).

44   *Thaler v. Vidal*, 43 F.4th 1207 (Fed. Cir. 2022).

45   *Thaler v. Perlmutter*, No. 1:22-cv-01564, 2023 U.S. Dist. LEXIS 145823, 3 (D.D.C. Aug. 18, 2023).

46   Daniel J. Gervais, *The Machine as Author*, 105 Iowa Law Review 2053, 2070 (2020).

A different but more intuitive starting point for a layperson might be that the *developer* of the generative AI system should own its creations. Indeed, pursuant to Sections 101 and 201(b) of the United States Copyright Act, works produced as "works made for hire" are those that are either "prepared within the scope of employment" or are "specially ordered or commissioned for use as a contribution to a collective work." Put differently, an "employer" possesses the copyright for a work crafted by their employee within the bounds of their "employment." Of course, addressing the issue of ownership of AI-generated works under this doctrine necessitates broadening the definition of "employment" (or a different non-employment work arrangement) to encompass the relationship between a programmer and their computer.[47]

The economic arguments around the implications of this extension are complex. On the one hand, one might imagine individual artists (or other creators) using an AI platform to create "digital twins" of themselves trained on their prior works (and that thus create in their individual style), asserting ownership over these AI systems, and (perhaps legitimately) then claiming ownership over the new works created by their individualized AI system as "fruits of their prior labor." Under such a scenario, the AI system complements the individual human creator. On the other hand, an AI platform allowed to train a model on all existing works across artists could create a system that significantly alters the need for human artists. A music label that owns a few recordings of a new artist may decide that it is economically more efficient to embed these into an AI system that replicates the artist rather than continuing their commercial relationship with the artist.

As we sort out the specific law and economics around the ownership of works created by one's "digital twin," it is also instructive to examine recent cases and discussion around joint production associated with generative AI creations. In February 2023, the USPTO concluded that a graphic novel comprising human-authored text combined with images generated by the AI service Midjourney constituted a copyrightable work, but that the individual images themselves could not be protected by copyright.[48] Recent guidelines issued by the U.S. Copy-

right Office mandate that human creators must specify, within the copyright registration application, the components originating from human authorship and those generated by an AI system.[49]

Delving a little deeper into the workings of generative AI makes it clear there are other scenarios that might be considered joint production. Most creations of a generative AI system are in response to a user "prompt." These prompts can often be extremely complex, incorporate a substantial number of examples or proprietary data (indeed, perhaps the most common way that pre-trained models like GPT-4 are being customized by corporations today uses a technique called "prompt tuning"), and the crafting of the "right" prompt could thus be thought of as a creative and productive exercise (again, today, there are specialists called "prompt engineers" who specialize in this kind of prompt creation, although it is possible that such prompt engineering will become less important in the future). Giving the creator of the prompt ownership over the AI-generated creation that emerges in response to their prompt could be seen as democratizing innovation, and in some settings, serving as an incentive for the party best positioned to create and introduce AI-generated works to the public. The argument supporting joint ownership of the copyright by both the programmer and the user responds to the reality that both parties played essential roles in creating the AI-generated work, and neither could have accomplished it independently.[50]

05
CONCLUSION

As we wait for greater clarity from either case law or legislative reform, the uncertainty around the use of copyrighted content for training purposes is being addressed partially in other ways. In July 2023, OpenAI signed a li-

---

47 Kalin Hristov, *Artificial Intelligence and the Copyright Dilemma*, 57 IDEA 431 (2017); Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5 (2012); Shlomit Yanisky-Ravid, *Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era: The Human-like Authors Are Already Here: A New Model*, 2017 MICH. ST. L. REV. 659 (2017).

48 See U.S. COPYRIGHT OFFICE, CANCELLATION DECISION RE: ZARYA OF THE DAWN (VAu001480196) (Feb. 21, 2023), 2, https://www.copyright.gov/docs/zarya-of-the-dawn.pdf.

49 The U.S. Copyright Office indicates it will evaluate the utilization of AI contributions case-by-case to ascertain whether the human element justifies copyright protection or not. For creators seeking copyright protection for works produced with the assistance of AI, it is crucial to keep track of the AI system's involvement in the final product. A more substantial degree of human involvement in the creative process enhances the likelihood of obtaining copyright protection; on the contrary, when an AI technology determines the expressive elements of its output, the generated material is not the product of human authorship and, therefore, is not copyrightable. COPYRIGHT REGISTRATION GUIDANCE: WORKS CONTAINING MATERIAL GENERATED BY ARTIFICIAL INTELLIGENCE, FEDERAL REGISTER 2023-05321 (Mar. 16, 2023).

50 Bruce E. Boyden, *Emergent works*, 39 COLUM. J.L. & ARTS 377, 384-388 (2015). Ralph D. Clifford, *Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand Up?*, 71 TUL. L. REV. 1675, 1695 (1997).

censing deal with the Associated Press to use its archive of news stories as training data. In August 2023, Google and Universal Music indicated they were negotiating a deal on how to license the voices and melodies of artists for AI-generated songs.[51] OpenAI has indicated that web sites can signal in their robots.txt file that they do not want their data used for training purpose (much like web sites could signal this to search engines), and a wide variety of "do not train" metadata standards and web sites are being created.

While these technological and market solutions may be a useful temporary fix as we rethink the boundaries of IP law, an excessively restrictive technological or licensing regime applied today could have the unintended consequences of favoring early movers (who have already gathered, organized and summarized large fractions of the world's existing information in their models' parameters) and may also favor larger players over smaller ones if, for example, licensing deals are bilateral and private rather than de jure.

In conclusion, today's dilemma is not unprecedented. Prior technological changes have necessitated de facto changes in how society regulates the allocation of IP rights and access rights between creators, consumers, and future innovators. Recall the debate around creating copies that occurred during the early days of the commercial Internet, a technology that challenged the act of creating a copy as the basis for assessing whether infringement had occurred. As Randall Davis pointed out: "...[A]ccessing digital information inevitably means making a copy, even if only an ephemeral copy. This copying action is deeply rooted in the way computers work. For example, when you view a page from the World Wide Web, several copies are made automatically — one so the document can be sent from the remote computer to your computer, a second when the document is loaded into memory, and yet another when it is displayed on the screen."[52] Clearly, considering each of these acts of copying infringements would challenge the economic viability of the Internet to the detriment of society. However, allowing unfettered replication and distribution via the Internet could have, in turn, drastically reduced the incentives to create.

We are at a similar turning point today. Granting excessive rights to creators and existing copyright holders will slow innovation in generative AI, especially for newer entrants, and could kill what may be a vibrant future industry of commercial AI-generated artistic content and other intellectual property. But granting no IP rights to a human over what they consider their highly individual artistic style or process of "creation," broadly defined, could have a chilling effect on both human creativity and broader societal innovation.

If left unchanged, it is more likely than not that the current IP regime will favor a dramatic shift away from human-led creation and towards one where more and more works are generated by machines. Thus, a key question society must consider is whether it is comfortable with a vast majority of future creation being done by machines rather than humans, and policy makers must in addition consider whether a future of primarily AI-generated creative works is on the optimal path. Consider that the abilities of systems like GPT-4 and PaLM are emergent — there is little consensus among experts on what the capabilities of generative AI will be in six months, let alone five or ten years from today, but there is consensus about the unpredictability of these capabilities. It is therefore unclear whether AI systems trained on AI-generated content will continue to "learn" new capabilities at the same rate (or at all), and it is entirely possible that they would be substantially better if there is a significant fraction of human-generated training data in the mix.

Finally, as humans, what will a creative conquest by AI do to our humanity? As John Seabrook concluded in his prescient 2019 New Yorker article about OpenAI's GPT-2, "One can imagine a kind of Joycean superauthor, capable of any style, turning out spine-tingling suspense novels, massively researched biographies, and nuanced analyses of the Israeli-Palestinian conflict. Humans would stop writing, or at least publishing, because all the readers would be captivated by the machines. What then?"[53] ■

> **"** *If left unchanged, it is more likely than not that the current IP regime will favor a dramatic shift away from human-led creation and towards one where more and more works are generated by machines*

51   Miles Kruppa, *How Frank Sinatra and Yo Gotti Are Influencing the Future of Music on YouTube*, Wall St. J. (Aug 21, 2023 9:00 am ET), https://www.wsj.com/tech/how-frank-sinatra-and-yo-gotti-are-influencing-the-future-of-music-on-youtube-971db915.

52   https://cacm.acm.org/magazines/2001/2/7457-the-digital-dilemma/fulltext.

53   John Seabrook, *Can a Machine Learn to Write for The New Yorker?*, New Yorker (Oct. 14, 2019), https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker.

CPI
SUBSCRIPTIONS

CPI reaches more than **35,000 readers** in over **150 countries** every day. Our online library houses over **23,000 papers**, articles and interviews.

Visit **competitionpolicyinternational.com** today to see our available plans and join CPI's global community of antitrust experts.

CPI COMPETITION POLICY® INTERNATIONAL