NORTH CAROLINA JOURNAL OF LAW & TECHNOLOGY
VOLUME 25, ISSUE 3: APRIL 2024

GENERATING DERIVATIVES: AI AND COPYRIGHT'S MOST TROUBLESOME RIGHT

Oren Bracha*

This Article examines broad arguments of infringing copyright's entitlement of the right of derivatives in the context of Generative AI ("GenAI") systems. Copyright owners make derivatives arguments against various activities in the GenAI supply chain even in the absence of substantially similar output. They make these arguments in an attempt to go around the limitations of the right of reproduction and establish liability even when reproduction arguments fail. The Article's argument is threefold. First, broad derivatives arguments fly in the face of copyright's basic subject matter and scope principles. This is because underlying such arguments is the least plausible and most normatively precarious conception of the right of derivatives. Second, courts can use two existing elements of the case law to reject these implausible claims: the requirement that to be a derivative, an infringing use must incorporate expression from the copyrighted work; and a firm application of the substantial similarity test of infringement. Third, the abuse of derivatives claims in the GenAI context exposes the more general difficulties associated with this right: a feeble normative basis, an ever-extending scope, and an obscure relation to the right of reproduction. These serious shortcomings should be remedied by a general, coherent conception of the derivatives right as a right of adaptations: the right of creating versions of the same work in its entirety, in a different expressive medium. Such a construal would allow a proper and normatively grounded application of the right both generally and in the specific context of GenAI.

* William C. Conner Chair in Law, The University of Texas School of Law.

TABLE OF CONTENTS

I. INTRODUCTION

The Generative AI ("GenAI") revolution is here. With ever-improving technology for low-cost generation of high-quality expressive works in a variety of media, it has arrived to the field of cultural production.[1] A disruptive technology if there ever was one,

[1] *See, e.g.*, Greg Bensinger, *Focus: ChatGPT launches boom in AI-written e-books on Amazon*, REUTERS (Feb. 21, 2023, 3:43 PM),

GenAI sends shock waves throughout the field. The technology holds much promise for satisfying demand for cultural materials and empowering creativity through hybrid human-machine models.[2] But the news is not all good. Many authors and copyright owners are worried about both uses of their work in the course of producing GenAI systems and the more general competitive threat posed by them.[3] Moreover, the specter of machines displacing humans from markets for cultural expression gives rise to deeper policy anxieties about the social effects of the technology.[4]

Many of these concerns and the disputes they spark find their way to copyright's gates. The result is a deluge of copyright infringement lawsuits against various entities involved in the GenAI production cycle.[5] In these legal actions copyright holders hurl at

---

https://www.reuters.com/technology/chatgpt-launches-boomI-ai-written-e-books-amazon-2023-02-21 []; Ian Tucker, *AI journalism is getting harder to tell from the old-fashioned, human-generated kind*, GUARDIAN (Apr. 30, 2023, 8:51 AM), https://www.theguardian.com/commentisfree/2023/apr/30/ai-journalism-is-getting-harder-to-tell-from-the-old-fashioned-human-generated-kind [https://perma.cc/F4SF-M3VQ]; Kevin Roose, *An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy*, N.Y. TIMES (Sept. 2, 2022), https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html [https://perma.cc/2M5X-4P9N].

[2] *See* Oren Bracha, The Work of Copyright in the Age of Machine Production 37 (Feb. 16, 2024) (unpublished manuscript) (on file with SSRN).

[3] *See, e.g.*, Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J. L. & ARTS 45, 77 (2017) (arguing that "[e]xpressive machine learning not only jeopardizes the market for the works on which it is trained, it also threatens to marginalize authors completely"); Pamela Samuelson, *Generative AI Meets Copyright*, 381 SCI. 158, 159 (2023) (observing that "[g]enerative AI seems poised to have substantial impacts on the careers of professional writers and artists"); Betsy Reed, *'ChatGPT Said I Did Not Exist': How Artists and Writers Are Fighting Back Against AI*, GUARDIAN (March 18, 2023, 12:00 PM), https://www.theguardian.com/technology/2023/mar/18/chatgpt-said-i-did-not-exist-how-artists-and-writers-are-fighting-back-against-ai [https://perma.cc/6TMU-65AY].

[4] *See* Bracha, *supra* note 2, at 37–41.

[5] At of the time of submitting this Article there were sixteen active GenAI copyright infringement lawsuits. Some of the more significant cases are: Andersen v. Stability AI Ltd., No. 3:23-cv-00201, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023) (a class action brought by artists against makers of AI image generators); Getty Images (US), Inc. v. Stability AI, Inc., No. 1:99-mc-09999, (D. Del. Feb. 2, 2023) (an action brought by Getty Images against the maker of an

defendants a broad variety of infringement claims. Some of the infringement arguments being asserted are within the traditional mold of copyright. Others, however, stretch the limits of the field and raise challenging doctrinal, conceptual, and policy questions. Ambitious arguments of the latter kind are fueled by the desire to extend broad liability upwards in the GenAI supply chain to reach actors with deeper pockets and effective control over the production and operation of the systems.

Typically, the first wave of attack by ambitious infringement arguments relies on copyright's most basic entitlement: the right to reproduce the copyrighted work.[6] Reproduction arguments take two forms: upstream arguments that focus on activities that happen early in the GenAI production cycle, such as making training copies or creating a model; and downstream arguments that focus on the generation of specific expressive output.[7] Both kinds of broad infringement arguments run into serious difficulties, especially in cases where there is no generated output that is substantially similar to any copyrighted work. In such cases, established principles and precedents tilt heavily against infringement, for two reasons.[8] First, the relevant activities are likely to be privileged as fair use.[9] Second, and more fundamentally, these activities are not within the domain of copyright because they involve only non-copyrightable subject

image generator AI for copyright infringement in millions of images in its collection); Does v. GitHub, Inc., No. 3:22-cv-06823, (N.D. Cal. May 11, 2023) (a class action brought against makers and distributor of an AI system for producing computer code); N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y Dec. 27, 2023) (an action by the New York Times against Open AI and others for using millions of its copyrighted news articles in the training of an AI model).

[6] 17 U.S.C. § 106(1).

[7] *See infra* Part III.

[8] *See id.*

[9] *See, e.g.*, Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 745 (2021); Pamela Samuelson, *Fair Use Defenses in Disruptive Technology Cases*, UCLA L. REV. (forthcoming 2024) (manuscript at 70) (on file with SSRN); Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. 1887 (2024).

matter, such as metainformation, high-abstraction structural elements, or mere physical aspects of works.[10]

Predictably, this is where the right of derivatives enters the stage and takes a central role in the unfolding GenAI copyright infringement drama.[11] Alongside the right of reproduction, copyright gives owners the right to prepare derivative works based on theirs.[12] Plaintiffs in GenAI cases turn to the right of derivatives in the hope of imposing the broad liability that ambitious reproduction arguments struggle to deliver. They use that right as a wild card: mobilizing it as a way around the subject matter and scope principles that limit the right of reproduction.[13] Plaintiffs are aided in this endeavor by some troubling features of the right of derivatives: its nebulous and potentially broad scope, its unclear rationale, and its mysterious relationship to the right of reproduction.[14]

The thesis of this Article is threefold. First, this Article argues that broad right of derivatives arguments fail for the same reasons that their reproduction counterparts do. These arguments try to apply the right of derivatives in ways that threaten basic copyright principles of subject matter (what informational elements are within the domain of copyright) and scope (the breadth of the right to exclude with respect to copyrightable subject matter). In doing so, the arguments deploy the least plausible and most normatively precarious interpretations of the right of derivatives, ones that must be rejected to maintain the integrity of the field. Second, copyright law already contains specific antidotes to such implausible readings of the right of derivatives: specific doctrinal requirements that, when applied by courts, keep tendencies for overbroad and implausible readings of the right in check. Third, the attempt to abuse the right

---

[10] *See* Bracha, *supra* note 2, at 25 (arguing that many broad GenAI infringement arguments fail on subject matter grounds).

[11] *See* Daniel J. Gervais, *AI Derivatives: The Application to the Derivative Work Right to Literary and Artistic Productions of AI Machines*, 53 SETON HALL L. REV. 1111 (2022) (predicting, accurately, that the right of derivatives would play a significant role in GenAI copyright infringement cases).

[12] 17 U.S.C. § 106(2).

[13] *See infra* Part V.

[14] *See infra* Part IV.

of derivatives in GenAI copyright cases exposes deeper and more systematic problems associated with this right. This Article argues that these problems should be addressed with a general remedy, rather than local fixes: construing the right of derivatives as a right of adaptations. The right of adaptation is the right to exclude others from uses that incorporate the entire work or a primary part thereof but recast it in a new medium.

The Article proceeds in six Parts. Part II briefly explains the technology of GenAI and its intersection with the production of cultural expression. Part III examines the broad right of reproduction GenAI infringement arguments—both upstream and downstream, their difficulties, and why they are likely to fail. Part IV introduces the right of derivatives and three of its most problematic features that are germane for the use and abuse of the right in GenAI cases: disagreements about the scope of the right, its precarious normative basis, and conceptual ambiguities associated with it. Part V surveys the universe of broad right of derivatives arguments in GenAI cases and critiques each. Part VI discusses the available solutions for the excesses of broad derivative right arguments: local doctrinal fixes already in existence in some strands of the case law, and, alternatively, a more general reading of the right as a right of adaptations. The latter, this Article argues, offers a coherent and systematic understanding of the right of derivatives that ameliorates its worst conceptual and normative drawbacks. The right of adaptations reading also points at the proper role that the right has to play in GenAI copyright infringement cases. Part VII concludes.

II. GENERATIVE AI AND COPYRIGHT INFRINGEMENT

This Part lays the foundations for explaining the turn to right of derivatives arguments in GenAI copyright infringement cases. It starts by briefly explaining the technology of GenAI and its relevance for copyrighted works. It then describes two categories of broad copyright infringement arguments that target two different segments of the GenAI supply chain: upstream and downstream arguments.

A. The Technology

The field of Artificial Intelligence ("AI") is concerned with building machines—today this usually means digital computers—that are capable of processing information, making decisions, and acting in ways that are rational.[15] Sometimes rationality is defined in this context as mimicking human intelligence.[16] Others prefer a broader concept of rationality, which seems apt in capturing both the ways in which the processes of machine intelligence may differ from human ones and the fact that in many areas intelligent machine capabilities have far surpassed human ones.[17] Machine learning is a dominant subfield of AI. Its distinctive feature is that the capabilities are achieved through a process of learning or training the system.[18] Finally Generative AI is a subset of AI, which relies on machine learning approaches. The distinctive trait of GenAI is its genericity: the system's ability to generate new and useful information goods.[19] An AI system that plays chess, drives a car, or analyzes weather patterns is not generative, but one that generates new protein sequences or texts is.

The GenAI production process is best understood as one involving the transition from concrete information goods to metainformation and then back to concrete information goods. The first transition is the *training* process. In training the system is

---

[15] STUART G. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 19 (4th ed. 2022).

[16] *See, e.g.*, *Artificial Intelligence*, IBM, (Dec. 6, 2022) https://www.ibm.com/design/ai/basics/ai/ [https://perma.cc/BZ5S-X73S] (defining AI as "[a]ny system capable of simulating human intelligence and thought processes").

[17] *See* Russell & Norvig, *supra* note 15, at 19–22 (discussing competing concepts of intelligence as fidelity to human performance or as more general rationality).

[18] ETIENNE BERNARD, INTRODUCTION TO MACHINE LEARNING 1 (2021).

[19] MOHAK AGARWAL, GENERATIVE AI FOR ENTREPRENEURS IN A HURRY 5 (2023) ("While traditional AI is designed to recognize or classify existing data, generative AI is able to generate novel and diverse outputs based on a given set of input parameters or conditions."). A somewhat more accurate definition of GenAI is as an AI system capable of generating new data of the same kind of that in it training set. *See* BERNARD, *supra* note 18, at 14 (describing generative modeling in AI as "the most difficult unsupervised learning task" which is "to learn how to generate examples that are similar to the training data").

exposed to a large amount of data relevant for its purpose, known as the *training set*. The purpose of training is to extract metadata: information about patterns and relations between elements of the data in the training set.[20] The metadata is embodied in a model: a complex set of parameters and "weights" that mathematically represent the extracted patterns.[21] A common example of such metadata is Large Language Models ("LLMs") that represent mathematically patterns and relations between basic elements of language.[22] At the other end of the process lies *generation* which uses the metadata in the model to create a new specific information good, hopefully one that matches the user's needs.[23] Generation is typically triggered by a prompt—some input from the user—that then initiates a process in which the system, by reference to the metadata in the model, constructs a new information good as output.[24] Within the field of expression, the process of GenAI can be applied to a growing array of media—including text, speech, images, videos, and music—resulting in an impressive capability of machine creation in these areas.

The GenAI production process is far more complex than the above distillation, both technologically and institutionally. It is best thought of as a supply chain which leads from the starting point of preexisting data to the end result of newly generated information goods.[25] This supply chain can be further broken down into discrete

---

[20] IAN GOODFELLOW ET AL., DEEP LEARNING 2–3 (2016) (describing machine learning as AI acquiring "their own knowledge, by extracting patterns from raw data").

[21] *See* BERNARD, *supra* note 18, at 81–83 (explaining the concept of a model).

[22] *See* RUSSELL & NORVIG, *supra* note 15, at 874–75 (discussing language models).

[23] *Id.* at 168–69 (explaining data generation).

[24] Stephen Wolfram, *What Is ChatGPT Doing . . . and Why Does It Work?*, STEPHEN WOLFRAM WRITINGS (Feb. 14 2023), https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/ [https://perma.cc/NS6D-2SEG] (explaining that the idea of GenAI such as ChatGPT is to start with a large data set and "[t]hen train a neural net to generate text that's 'like this'. And in particular, make it able to start from a 'prompt' and then continue with text that's 'like what it's been trained with' ").

[25] Katherine Lee et al., *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, J. COPYRIGHT SOC'Y U.S.A. (forthcoming 2024) (manuscript at 29) (on file with SSRN).

technical stages.[26] On the institutional side, typically different entities would be involved in different ways in the relevant actions within each of the stages.[27] Rarely would it be the same entity who assembles the data in the training set, conducts the training, deploys the system, and provides the user's prompt to trigger the generation. Moreover, GenAI technology can be deployed through a wide variety of models such as syndicated services to users, embedding in software products, or seamless incorporation into other systems and products.[28] The result is further diversity and complexity of the entities involved and their mode of interaction with the GenAI process.

B. GenAI and Copyright

GenAI has many valuable uses in a variety of fields. However, the relevance and power of technology in the specific area of cultural expression is apparent. GenAI can cost-efficiently generate high-quality information goods in a growing variety of media ranging from image to music, and almost anything in between. As these systems are being broadly deployed, the effects on the field of cultural production are likely to be profound. Many of those effects are positive. In an expanding array of media, technology can supply consumer demand at low cost and on an impressive level of customization and quality. You want a digital image of an evil black cat dressed as a Roman Emperor? Just ask DALL-E to generate one for you. You happen to fancy a haiku about aliens coming from another dimension? ChatGPT can help. Furthermore, the potential is not limited to satisfying consumptive demand. GenAI holds a significant promise for empowering human creativity, either by being incorporated into creative productivity tools or through various hybrid models of human-machine creativity.[29]

But not all news is good. The disruptive effect of GenAI within the field of cultural production also triggers various concerns among creators and right owners, as well as observers more generally

---

[26] *Id.* (manuscript at 31) (analytically breaking down the GenAI supply chain into eight stages).

[27] *Id.* (manuscript at 29).

[28] *Id.*

[29] *See* Bracha, *supra* note 2, at 14.

preoccupied with cultural policy. The range is wide. On one end of the spectrum are traditional threats of corrosive appropriation of copyrighted works. A GenAI system that generates a digital image almost identical to a copyrighted picture that was present in its training set obviously raises concerns of copyright infringement and adverse effects on creators' incentive and fair compensation, both being purposes that copyright is meant to safeguard.[30] These threats are amplified by the power and ubiquity of the technology. A different fear, that is often not sufficiently distinguished from the previous one, is that of effective competition from GenAI generation of new and different materials that can serve as good substitutes for those created by humans.[31] Appropriation of specific expression aside, GenAI production may erode the ability of creators to extract compensation for their works in the market by providing low-cost, high quality independent substitutes. Further on the scale are concerns of general cultural and social policy that arise from a prediction that GenAI will come to dominate production in certain markets for expression. In markets where such predictions materialize GenAI production might displace human creators, resulting in troubling effects on livelihoods in creative industries, opportunities for access to the inherent value of creative activity, and the potential for paradigm-breaking cultural innovation.[32]

Because copyright is our main tool for dispensing cultural policy, the entire range of concerns over adverse GenAI effects is laid at its doorstep. This takes the form of a deluge of legal actions in which various stakeholders assert claims for copyright infringement against entities involved in the GenAI supply-chain. The range of infringement arguments and theories deployed in these lawsuits is broad. Some of these arguments are firmly within copyright's traditional contours. When a GenAI system generates an

---

[30] *See* Sobel, *supra* note 3, at 65 (observing that GenAI "output may infringe copyright in the pre-existing work or works to which it is similar.").

[31] *See* Benjamin Sobel, *A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning*, *in* A.I & INTELL. PROP. 19 (Reto Hilty et al., eds. 2021) (arguing that "[s]ome uses of machine learning are 'market-encroaching:' these uses of AI plausibly threaten the market for the copyrighted works that comprise their training data").

[32] *See* Bracha, *supra* note 2, at 37.

image or text that bears close similarity to a copyrighted work in the training set, there is a strong case for infringement. Such cases of substantially similar generated output may raise complex questions about the agent to whom the infringing act can be imputed and the extent to which liability can be attributed, directly or indirectly, to different entities in the GenAI supply chain. But the basic claim for infringement is a fairly strong and conventional one.

Plaintiffs, however, are often interested in expanding liability to entities other than those captured by conventional infringement claims. Several vectors converge to create this motivation: attempts to get legal leverage against entities with deeper pockets and stronger control in the GenAI supply chain, interest in curbing competition from independent GenAI generated goods or at least get a share of their value, and even strategies designed to slow down the rise of GenAI technology in markets for expression as a remedy to general cultural policy concerns.[33] The result is the deployment of less conventional and more ambitious arguments of copyright infringement. Ambitious infringement arguments come in many forms and flavors. They fall, however, into two groups: upstream and downstream arguments.

Upstream arguments target activities and entities located at the early stages of the GenAI supply chain. Typically, these arguments would target either the making of training copies—reproduction of copyrighted works in the training set strictly for purposes of training, or the model—the metadata extracted from the specific works in the training set. The utility of upstream arguments for plaintiffs is apparent. If successful, these arguments strike at the heart of the GenAI supply chain: they extend liability to entities with central control over producing the systems, potentially with no regard to the question of whether the relevant system generates substantially similar output at all.

Ambitious downstream arguments focus on the generated output. They differ from more conventional infringement claims by trying to extend liability to generated output whose level of similarity to the relevant copyrighted works is limited. The strategic value of this type of argument is achieved in two stages. First, broad

[33] *Id.*

downstream arguments attempt to significantly expand the number of cases in which generated output is subject to liability. Second, the hope is to extend this expanded control upward, by asserting that upstream entities, rather than just those located closer to the generation stage, are liable for the infringing output, either directly or under one of copyright's secondary liability doctrines.

III. RIGHT OF REPRODUCTION ARGUMENTS AND THEIR LIMITS

The main route for deploying broad infringement arguments is through copyright's most basic right: the right of making copies, also known as the right of reproduction.[34] However, as explained below, broad right of reproduction arguments run into serious difficulties, in either the upstream or downstream version. It is the limits of the right of reproduction that eventually lead copyright owners to turn to the right of derivatives for the rescue, in the hope of achieving through that route the broad infringement scope that the right of reproduction fails to deliver.

A. Upstream

There are two kinds of upstream reproduction arguments: one targets training copies, the other focuses on the model. Each argument runs into serious difficulties and, therefore, has uncertain prospects of succeeding.

1. Training Copies

The training copies argument is simple. To train a model—that is, to extract the metadata represented by it—digital copies of the informational items in the data set must be made. Computers—at least under current technology—simply cannot process the data and extract from it the relevant information in the absence of such copies. Often these training copies are made by "scraping" information that is publicly accessible via the internet, but even other sources of training data require the making of copies. These physical copies are squarely within the statutory definition of reproducing a work.[35] And since the copies are literal—they follow

[34] 17 U.S.C. § 106(1).

[35] *See id*. (giving the owner the exclusive right to "reproduce the copyrighted work in copies"); *see also* 17 U.S.C. § 101 (defining "copies" as "material objects

exactly the patterns of the original—copyright's infringement test and its requirement of substantial similarity are necessarily satisfied.[36] The result: to the extent the work reproduced in the training copy is under copyright, the right of reproduction is infringed.[37]

Note the strength of the reproduction argument. Its logic rests the entire weight of the argument on the act of reproduction, the physical act of embodying identical informational patterns in a physical object. As a result, infringement occurs at the moment of such embodiment, no matter what happens next. Reproduction is reproduction whether or not it subsequently leads to generation of any expression similar to the protected work, or indeed to any additional exposure by anyone to the expressive content of the work. Thus, infringement occurs irrespective of any similar output, or even whether the system has any expressive output at all. In its strongest version, the argument sweeps in a very broad range of training reproduction of copyrighted works, even for non-generative purposes, such as systems for car driving or facial recognition trained on copyrighted images.[38]

While deceptively simple and powerful, the training copies argument quickly runs into serious difficulties. A firm line of precedents casts a deep doubt on the outcome of infringement in cases of non-expressive copies—physical reproduction that does not involve further human exposure to the expressive content of the work.[39] These precedents concede that non-expressive copies

---

. . . in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device").

[36] *See* Arnstein v. Porter, 154 F.2d 464 (2d Cir. 1946).

[37] *See* Sobel, *supra* note 3, at 61.

[38] *See* Lemley & Casey, *supra* note 9, at 745.

[39] *See*, *e.g.*, Sega Enters. v. Accolade, Inc., 977 F.2d 1510, 1514 (9th Cir. 1992); Sony Computer Entm't, Inc. v. Connectix Corp., 203 F.3d 549, 603 (9th Cir. 2000); A.V. v. iParadigms, LLC, 562 F.3d 630, 645 (4th Cir. 2009); Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014); Authors Guild, Inc. v. Google, Inc., 804 F.3d 202 (2d. Cir. 2015). Mathew Sag has appropriately dubbed such cases "nonexpressive" uses. Matthew Sag, *Copyright and Copy-reliant Technology*, 103 NW. L. REV. 1607, 1068 (2009). A large number of commentators follow these precedents to conclude that in the absence of substantially similar output training

constitute prima facie infringement of the right of reproduction, but insist on exempting it as fair use. The fountainhead of this line of decisions is *Sega v. Accolade*.[40] In *Sega*, the court found that intermediate copying of copyrighted computer code—copying that was designed only to extract unprotectable information about communication protocols of the Sega game console—presumptively violated the reproduction right.[41] However, it went on to find the reproduction to be fair use.[42] The court's reasoning emphasized the non-appropriative character of the copying—the fact that the copying was only done to extract non-protectable information without exposing anyone to the protected expression—and the public benefit of using the extracted information to independently create new computer games and increase variety and competition in that market.[43] *Sega* became the foundation on which later cases, most notably those involving background digital reproductions of copyrighted books for purposes of the Google Books service, consistently exempted non-expressive copies as fair use.[44]

Facing this significant hurdle, those who claim training copies infringement try to distinguish the fair use precedents with a variety of legal and policy arguments.[45] One such argument is that the basis

---

copies are exempted as fair use. *See, e.g.*, Lemley & Casey, *supra* note 9, at 745; Pamela Samuelson, *Generative AI Meets Copyright*, 381 SCI. 158, 159–61 (2023) (surveying the possible fair use analysis of claims against training copies in ongoing lawsuits); Samuelson, *supra* note 9 (manuscript at 55); Sag, *supra* note 9, at 24–30; Enrico Bonadio et al., *Can Artificial Intelligence Infringe Copyright? Some Reflections*, *in* RSCH. HANDBOOK ON INTELL. PROP. & A.I. 247 (Ryan Abbott ed., 2022); Jessica L. Gillotte, *Copyright Infringement in AI-Generated Artworks*, 53 U.C. DAVIS L. REV. 2655, 2680 (2020); Daryl Lim, *AP & IP Innovation: Creativity in An Age of Accelerated Change*, 52 AKRON L. REV. 813, 847 (2018); James Grimmelmann, *Copyright for Literate Robots,* 101 IOWA L. REV. 657, 661–65 (2016).

[40] Sega Enters. Ltd. V. Accolade, Inc., 977 F.2d 1510, 1514 (9th Cir. 1992).

[41] *Id.* at 1518.

[42] *Id.* at 1522–28.

[43] *Id.*

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

[45] *See* Sobel, *supra* note 3, at 77–79 (arguing that "market-encroaching" uses may not be entitled to the fair use privilege even if the competing work does not

of the fair use analysis disappears in cases where the non-expressive copies facilitate further reproduction down the road that enables access to the expression of the protected work.[46] Whatever the merit of this argument, it has no bearing on the set of cases discussed here, which by stipulation are exactly those where no substantially similar expression is generated. For the same reason, even if it succeeded, the distinction would radically limit the breadth of the training copies argument: rather than the gigantic universe of all training reproduction irrespective of downstream activity, it would only apply to the much smaller subset of training reproduction that leads to generated similar expression.

Another possible distinction points to competition as such. GenAI non-expressive copies are different, the argument goes, because they are "market encroaching."[47] While the training copies themselves do not compete with the copied works or divert some of their market demand, they facilitate the production of a GenAI system whose output of independent works does create demand-diverting competition. Unfortunately, this distinction runs against both precedent and fundamental copyright principles. In *Sega*, the purpose of the non-expressive intermediary copies was exactly to extract unprotected metainformation that then facilitated the creation of independent computer games that competed with those of the copyright owner.[48] Not only did the court not rule that this denied the fair use privilege to the user, but it also found this kind of competition to be in the public interest and, therefore, in strong support of the fair use conclusion.[49]

---

copy expression); Andrew W. Torrance & Bill Tomlinson, *Training is Everything: Artificial Intelligence, Copyright, and "Fair Training"*, 128 DICK. L. REV. 233 (2023) (surveying arguments that training copies are not fair use); Mathew Sag, *Copyright Safety for Generative AI*, 61 HOUS. L. REV. 295, 308–10 (2023) (critically discussing attempts to distinguish GenAI).

[46] *See* Sag, *supra* note 45, at 312–13.

[47] See Sobel, supra note 3, at 77–79; Sobel, supra note 31, at 231.

[48] *Sega*, 977 F.2d at 1523.

[49] *Id.* (explaining that the "increase in the number of independently designed video game programs . . ." created by defendant by extracting non-protectable information from plaintiff's games "is precisely this growth in creative expression, based on the dissemination of other creative works and the

This analysis expresses a fundamental distinction in copyright law between appropriative and independent competition. Copyright does not condemn any competition to copyrighted works. Its purpose is to prevent competition based on appropriation, namely, diverting demand for a copyrighted work by creating a substitute for it via copying, thereby enabling the copier to incur a much lower cost of production and thus undermining the market for the original.[50] Competition based on independently creating substitutes, even if deemed "harmful" from the private point of view of a copyright owner, is not a cognizant harm from the point of view of copyright law. Rather, independent competition is a desirable effect, in line with promoting copyright's basic purpose of innovation in the cultural sphere.[51] For this very reason, copyright law permits extracting a variety of unprotectable elements from copyrighted works, including the knowledge of how to produce such works, to then subsequently create independently competing works. For the same reason, when the extraction of the unprotected information is done via a non-expressive copy, the resulting competing expressive product is seen as a boon rather than a danger, one that supports rather than undermines a finding of fair use.[52]

Moreover, as this Author argues elsewhere, reproduction in non-expressive training copies is non-infringing because of a more fundamental reason than fair use.[53] It is non-infringing because such

---

unprotected ideas contained in those works, that the Copyright Act was intended to promote").

[50] *See* Oren Bracha & Talha Syed, *Beyond the Incentive–Access Paradigm? Product Differentiation & Copyright Revisited*, 92 TEX. L. REV. 1841, 1849 (2014) (describing the copyright policy problem as based on nonexcludability and the gap between the costs of innovation and imitation).

[51] U.S. CONST. art. I, § 8, cl. 8 (mandating that Congress's power to legislate in the area of copyright and patents is granted "to promote the Progress of Science and useful Arts").

[52] *See* Samuelson *supra* note 9 (manuscript at 66) (discussing "precedents upholding fair use defenses when defendants' ultimately non-infringing products compete with the plaintiff's products").

[53] *See* Bracha, *supra* note 2, at 25; *see also* Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem,* 93 WASH. L. REV. 579, 595–96 (2018) (focusing mainly on the fair use analysis, but observing that "[c]ourts have also yet to confront whether unauthorized copies

reproduction does not pertain to any copyrightable subject matter.[54] The domain of copyright is limited to expression.[55] It does not encompass a variety of other informational elements, no matter how valuable, including: the informational content of works, high-abstraction level structural expressive elements, stock elements within a particular genre, and metainformation about the work.[56] The latter category of non-copyrightable elements includes metainformation that pertains to skills and techniques of producing similar works.[57] If someone reads a copyrighted book, learns how to write similar books from it, and produces such a work without appropriating any of the original's expression, that person has not taken any copyrightable subject-matter.[58] Non-expressive copies in general and GenAI training copies in particular involve the exact same circumstance; the only appropriated information is unprotected metainformation.

Some are likely to object that non-expressive reproduction is different: unlike cases of mere extraction of metainformation, training copies necessarily involve reproduction of the exact physical patterns that represent the *expression* of a copyrighted work.[59] But this would be falling into the trap of the physicalism fallacy.[60] Copyright as a field whose domain is expression does not

made for training AI are necessarily infringing copies" and suggesting other possible reasons for non-infringement).

[54] Bracha, *supra* note 2, at 25.

[55] 17 U.S.C. § 102(a) (providing that copyright subsists in "works of authorship").

[56] 17 U.S.C. § 102(b) (delineating uncopyrightable subject matter).

[57] The principle goes back to the foundational period of modern copyright in eighteenth century England. *See* Millar v. Taylor (1769) 98 Eng Rep. 216 (KB) (explaining that copyright leaves "all the knowledge that can be acquired from a contents of a book . . . free for every man's use," whether that knowledge is "mathematics, physic, husbandry").

[58] *See id.* ("[I]f, reading an epic poem, a man learns to make epic poems of his own; he is at liberty.").

[59] *See* Sobel, *supra* note 3, at 73–74 (arguing that "[n]o human can rebut an infringement claim merely by showing that he has learned by consuming the works he copied," and therefore "[i]f future productivity is no defense for unauthorized human consumption, it should not excuse robotic consumption, either").

[60] *See* Bracha, *supra* note 2, at 23.

care about mere physical acts. Physical acts, such as reproduction or distribution, are significant within copyright only to the extent they facilitate access to a work's *expressive value*.[61] Non expressive reproduction is simply an irrelevant physical fact that has no relevance for copyright. The result: non-expressive reproduction—including that of training copies—should be filtered out as the kind of copying that does not involve any taking of copyrighted expression.[62] Thus, non-expressive training copies do not infringe on subject matter grounds, long before reaching the back-end question of fair-use.

2. The Model

As an alternative to training-copies, upstream reproduction arguments may target the GenAI model. The argument here is that the model itself—the mathematical values representing patterns and relations extracted from multiple works—is a form of reproduction of these works and, therefore, infringing. This reproduction argument is even weaker than the training-copies one. The model is not a reproduction of the informational patterns of any particular work, but simply a representation of the aggregate metainformation about many works and the relations between them. This metainformation does not embody the expression of a specific work any more than a spreadsheet containing information about the frequency of the appearance of certain phrases in a group of texts replicates the expression in any of those texts. Consequently, the model is a form of unprotectable metainformation whose creation does not involve any copyrightable subject matter.

---

[61] *Id.* at 23.

[62] *See e.g.*, Mattel, Inc. v. MGA Ent., Inc., 616 F.3d 904, 915 (9th Cir. 2010) (referring to the need to "filter out the unprotectable elements" in analyzing infringement); Kohus v. Mariol, 328 F.3d 855 (6th Cir. 2003) (explaining that "[t]he essence of the first step [in analyzing improper appropriation] is to filter out the unoriginal, unprotectible elements"); Stromback v. New Line Cinema, 384 F.3d 283, 296 (6th Cir. 2006) (discussing filtering of unprotected elements); Blehm v. Jacobs, 702 F.3d 1193 n.4 (10th Cir. 2012) ("We . . . filter out unprotected elements from the author's protected expression."); Rentmeester v. Nike, Inc., 883 F.3d 1111, 1118 (9th Cir. 2018) ("Before that comparison can be made, the court must 'filter out' the unprotectable elements of the plaintiff's work.").

One may object by arguing that, at least in cases when the specific works can be reconstructed from the metainformation in the model—a process that is sometimes referred to as "memorization"—the metainformation is a form of reproducing the expression itself.[63] As long as the relevant information that enables the system to reconstruct a work in the training set is "conveyed" by the metainformation in the model, this is the functional equivalent of reproduction and there is no reason to insist on direct one-to-one mapping of the informational patterns of expression.[64] Moreover, other forms of reproduction that operate by packaging expressive information into metadata that can be converted back into expression, are broadly seen as reproduction of expression. This is the case with an MP3 file with respect to music, other digital compression standards, or simply a literary work that is written in code and comes with a key.[65] However, this analogy fails. Even in cases where a work is reconstructed from the model, there remains a significant difference—one which is crucial for copyright purposes—between metadata representation of a specific work, as in the case of the MP3 file, and aggregate metadata about a large group of works, as in a GenAI model.[66] More importantly, for current purposes, by stipulation, we are dealing only with cases that do not involve output that is substantially similar to any specific work in the training set. In such circumstances it is beyond doubt that, as far as copyright law is concerned, the model is not a reproduction of the expression of any particular work, but simply non-expressive metainformation about many works. As such,

---

[63] *See* Sag, *supra* note 45, at 312 (arguing that when models " 'memorize' particular works in the training data" this is likely infringing reproduction); Sobel, *supra* note 3, at 64 (arguing that when machine models are "overfitted," they constitute infringement).

[64] Sag, *supra* note 45, at 312; Sobel, *supra* note 3, at 64.

[65] UMG Recordings, Inc. v. MP3.com, Inc., 92 F. Supp. 2d 349, 350 (S.D.N.Y. 2000) (finding reproduction of music in MP3 format to be infringing).

[66] The difference is that the aggregate metadata can be put to many uses, many of which are not infringing, while coded representation of one work is a standardized means of communicating that specific work. *See also* Sobel, *supra* note 3, at 64 (observing that "[i]f a trained model *always ends up replicating its input data*, it would be sensible to call the model itself a copy or a derivative work") (emphasis added).

creating a model does not involve the reproduction of any copyrightable subject matter.

B. Downstream

On the downstream side, broad infringement arguments shift their focus to generated output. They attempt to capture cases where such output bears only remote or diffused similarity to work in the training set. The strategy here for achieving broad control of AI systems is two-tiered. In the first stage, broad downstream arguments increase the reach of liability for generated output beyond the standard scope of copyright. In the second stage, the increased scope is leveraged to implicate in the infringing acts actors who control crucial phases of the GenAI supply chain but are located at a relatively remote point from the generation stage.

This maneuver of extending liability upstream for downstream activities can be accomplished in various ways. One route is to argue that the upstream actor—for example, the entity who produced training copies or trained the model—by virtue of creating the system, is the agent directly responsible for the downstream activity that produced the output.[67] A second route is to argue that downstream actors' infringing activity may deprive upstream activity of the characteristics that led to the conclusion that it was not infringing. Most commonly, the argument here is that producing non-expressive copies upstream stops being fair use when the activity facilitates the production of a system used to produce expressive output downstream similar to works that were reproduced in the training copies.[68] A third alternative is to concede

[67] *See e.g.*, Complaint at 13, Andersen v. Stability AI, Ltd., No. 3:23-cv-00201-WHO, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023) (arguing that defendant "uses the compressed copies in generating its output in response to Text Prompts").

[68] *See e.g.*, Complaint at 13, Getty Images (U.S.), Inc. v. Stability AI, No. 1:23-cv-00135-GBW (D. Del. Feb. 3, 2023) (arguing that the unauthorized copies of plaintiff's images "were made with the express aim of enabling Stability AI to supplant Getty Images as a source of creative visual imagery"); Complaint at 31, *Andersen v. Stability AI, Ltd.*, 2023 WL 7132064 (arguing that "Defendants are using copies of the Training Images interconnected with their AI Image Products to generate digital images and other output that are derived exclusively from the Training Images, and that add nothing new. . . . that act as market substitutes for the underlying Training Images"). *See also*, Sag, *supra* note 45, at 312 (arguing

that only downstream actors have direct liability for infringing downstream reproduction, but then use broad construal of such liability as an important element for imposing secondary liability on upstream actors.[69] The basic structure of all three routes is the same: expand downstream liability for output, and then leverage it to impose upstream liability for other activities in the GenAI production process.

There are many variants of broad infringement arguments for generated output, but the most common and representative one is the argument of infringement by appropriating a creator's "style." While ambiguous in its meaning and often relying on other causes of action, such as unfair competition or right of publicity, there is a dominant copyright infringement version of the argument.[70] This argument holds that generated output bears a sufficient level of similarity, not to any particular copyrighted work, but to an aggregate of structural elements that are characteristic of the corpus of work of a specific creator.[71] As one artist framed the argument, "I can see my hand in [it], but it's not my work."[72]

Similar to upstream arguments, broad downstream arguments clash with fundamental copyright principles and, therefore, run into serious difficulties. In the case of copying style variants, the fatal flaw is twofold. First, copyright applies to works of authorship: well defined expressive bundles such as a novel, a painting, or a sculpture.[73] It does not apply to fabricated expressive entities that conceptually collect elements from different independent works,

---

that "[i]f ordinary and foreseeable uses of generative AI result in model outputs that would infringe on the inputs . . . then the nonexpressive use rationale [for fair use] would no longer apply").

[69] *See e.g.*, Complaint at 32, *Andersen v. Stability AI, Ltd.*, 2023 WL 7132064 (arguing that defendant is liable for vicarious copyright infringement because it allowed others to "cause the AI Image Product to rely more heavily on… [an] artist's prior works to create images that can pass as original works by that artist").

[70] *See e.g.*, Complaint 30–33, *Andersen v. Stability AI, Ltd.*, 2023 WL 7132064.

[71] *See* Bracha *supra* note 2, at 32.

[72] Kelly Mckernan, *Artists vs. AI*, NPR (Jan. 30, 2023, 6:32 PM), https://www.npr.org/2023/01/30/1152653269/artists-vs-ai [https://perma.cc/NTF3-GHCL].

[73] 17 U.S.C. § 102(a).

such as styles.[74] Therefore, the argument of copying style is made with respect to a unit of analysis that copyright law does not recognize. Second, style in the sense of common, structural expressive elements is exactly the kind of high-abstraction subject matter that established precedents insist is outside the domain of copyright and therefore designate as unprotectable "ideas."[75] Unprotectable high-abstraction elements, such as style, are filtered out as part of copyright's infringement test and cannot be the basis for establishing the substantial similarity that is required to satisfy it.[76] Finally, even if other variants of broad downstream arguments avoid the unit of analysis and subject matter problems, they are still very likely to crash against copyright infringement test's requirement of substantial similarity. What makes these arguments broad is exactly the fact that they target remote levels of similarity between the generated and the copyrighted work, which in turn makes establishing substantial similarity less likely.[77]

C. Taking Stock

In sum, broad arguments for the infringement of the reproduction right that try to extend copyright liability to actors who control central stages of the GenAI production process face considerable difficulties, in either their upstream or downstream version. This is particularly true in cases where the GenAI system does not generate output that is closely similar to particular copyrighted works in its training set. Upstream arguments, targeting either training copies or the model, try to capture reproduction which is outside copyright's domain because it does not involve any copyrightable subject matter. Moreover, even if successful in clearing the subject matter hurdle, these arguments face a firm line

---

[74] *See* Bracha, *supra* note 2, at 33–34.

[75] Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930); Dave Grossman Designs, Inc. v. Bortin, 347 F. Supp. 1150, 1156, 174 U.S.P.Q. 217 (N.D. Ill. 1972) (observing that "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").

[76] *See* cases cited *supra* note 62.

[77] *See* Bracha, *supra* note 2, at 33.

of precedents that exempt the making of non-expressive copies as fair use. Downstream reproduction arguments, targeting low-similarity generated output, have their own share of woes. Some versions of these arguments try to attach protection to contrived informational goods not recognized by copyright as a unit of protection. More broadly, these arguments are likely to fail the infringement test either because copied elements are unprotectable or because they are not substantially similar.

IV. THE RIGHT OF DERIVATIVES

Enter the right of derivatives.[78] Faced with the considerable difficulties in grounding broad infringement arguments in the right of reproduction, plaintiffs in GenAI cases often turn to the right of derivatives.[79] Here, the right of derivatives performs the function of a wildcard: the task appointed to it is to suspend or at least go around the limiting features of the right of reproduction, thereby securing the broad liability that reproduction arguments cannot. The origin of this dynamic is rooted in the various difficulties associated with the right of derivatives and its obscure relationship with the right of reproduction. This Part explains the right of derivatives and some of its most problematic features as the basis for properly analyzing its application to GenAI infringement arguments.

The right of derivatives was officially added to the Copyright Act in 1976.[80] Its addition was the culmination of an over-a-century-long intellectual development in the course of which the dominant understanding of the scope of copyright had transformed.[81] The mid-nineteenth century conception of copyright revolved around a narrow understanding of the field as literally

---

[78] 17 U.S.C. § 106(2).

[79] *See infra* Parts V.A.1, V.B.1.

[80] 17 U.S.C. § 106(2).

[81] OREN BRACHA, OWNING IDEAS: THE INTELLECTUAL ORIGINS OF AMERICAN INTELLECTUAL PROPERTY, 1790-1909, at 146–87 (2016). Daniel Gervais, *The Derivative Right, or Why Copyright Law Protects Foxes Better than Hedgehogs*, 15 VAND. J. ENT. & TECH. L. 785, 792–96 (2013); Pamela Samuelson, *The Quest for a Sound Conception of Copyright's Derivative Work Right*, 101 GEO. L.J. 1505, 1511–17 (2013).

"copy-right."[82] At the heart of this understanding stood the strict idea that a "copy" was a close replica of the protected work, to the extent of often excluding from the scope of the right even mildly more remote secondary uses, such as translations or dramatizations.[83] Gradually, driven by a rising conviction that creators should be able to internalize the full social value of the use of their work in any form, the scope of the right was expanded to include a greater range of secondary uses and weaker degrees of similarity.[84] In 1976, this process crystallized into Congress' addition of a new right given to owners, alongside the right of reproduction.[85] The new entitlement conferred on owners the exclusive right to "prepare derivative works based upon the copyrighted work."[86] The definition of a derivative work consists of a core description of the concept as "a work based upon one or more preexisting works," followed by a list of specific examples, as well as the inclusion of "any other form in which a work may be recast, transformed, or adapted."[87]

The 1976 express statutory addition of the right of derivatives was clearly meant to expand the scope of copyright. However, it also resulted in a long series of ambiguities and difficulties pertaining to the new right. The right of derivatives is copyright's most troublesome entitlement because of its obscurity and precarious policy rationale.[88] The debates, as well as the doctrinal and interpretive ambiguities surrounding the right, are many.[89]

---

[82] *See, e.g.*, Emerson v. Davies, 8 F. Cas. 615, 621 (C.C.D. Mass. 1845) (No. 4436).

[83] BRACHA, *supra* note 81, at 146–58.

[84] *See* Gervais, *supra* note 81, at 792–96; Oren Bracha, *Before an Image Was Worth A Thousand Words: Ben-Hur and Copyright's Right of Derivatives*, *in* CIRCULATION AND CONTROL: ARTISTIC CULTURE AND INTELLECTUAL PROPERTY IN THE NINETEENTH CENTURY 195 (Marie-Stéphanie Delamaire & Will Slauter eds., 2021).

[85] 17 U.S.C. § 106(2).

[86] *Id.*

[87] 17 U.S.C. § 101.

[88] *See* WILLIAM F. PATRY, PATRY ON COPYRIGHT § 3:46 (2024) (observing that "the principles governing derivative works" are the "'most troublesome' doctrine in copyright").

[89] *See* Samuelson, *supra* note 81, at 1510 & n.27 (observing that "[m]ysteries abound about the proper scope of the derivative work right").

Explained below are the three problematic dimensions of the right of derivatives that are most germane for understanding its application to GenAI infringement arguments. These include deep difficulties related to the rights' scope; basic rationale, if any; and its relation to the right of reproduction.

A. Scope

It is hard to dispute that the purpose of the right of reproduction is to expand the scope of copyright law's right to exclude by including at least some activities or uses not captured by the right of reproduction.[90] There is some disagreement, however, on what exactly is encompassed by this expanded scope of the right of derivatives and how broadly it expands.

Some commentators advocate for a broad conception of the right of derivatives that encompasses every possible valuable market for any secondary use of the work. Professor Paul Goldstein argues that the role of the right of derivatives is to "enable the copyright owner to proportion its investment to the level of expected returns from all markets."[91] Many courts have followed suit and applied the right of derivatives in a loose, freewheeling fashion. The thrust of these cases is to extend the right to any imaginable secondary market of the work, including trivia question games based on a television show,[92] still photographs of a performance of a choreographic work,[93] a real-life replica of the Batmobile as a comic character,[94] and computer chips that accelerate computer games.[95] These decisions, some courts' loose approach to the derivatives category,

---

[90] Otherwise, the right would be completely superfluous. That the right is superfluous because it only applies to situations already covered by reproduction and other rights is exactly the conclusion of a leading treatise writer. *See* MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 8.09[A][1] (2018) (arguing that the derivatives right is superfluous).

[91] Paul Goldstein, *Derivative Rights and Derivative Works in Copyright*, 30 J. COPY. SOC'Y U.S.A. 209, 227 (1983).

[92] *See* Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc., 150 F.3d 132 (2d Cir. 1998).

[93] Horgan v. MacMillan, Inc., 789 F.2d 157, 162–63 (2d Cir. 1986).

[94] DC Comics v. Towle, 802 F.3d 1012, 1026 (9th Cir. 2015).

[95] Midway Mfg. Co. v. Artic Int'l, Inc., 704 F.2d 1009, 1013 (7th Cir. 1983).

and the underlying principle of capturing every secondary market appear to make the right of derivatives almost limitless.

By contrast, most commentators, troubled by the implications of a broad right of derivatives, advocate narrow readings or even abolition of this right.[96] The motivation of these commentators is usually concern about the precarious normative justification of the right or its troubling incongruity with copyright's general principles and purposes. The result is various proposals to circumscribe the derivatives right and, consequently, the universe of secondary uses of copyrighted works to which the right applies. Some proposals argue for a narrow reading of the right by reference to the specific uses expressly listed in the definition of the term "derivative work."[97] Another lists principles to inform the application of the right, designed to serve as "notes of caution."[98] Yet others, suggest an amended definition that narrows down the sphere of derivatives to those that exhibit minimal originality, have significant effect on the economic feasibility of creating the primary work, or, by reference to some other element.[99] Finally, various commentators advocate an array of other mitigating mechanisms, ranging from a compulsory license scheme to a limited remedy.[100]

---

[96] *See generally* Derek E. Bambauer, *Faulty Math: The Economics of Legalizing the Grey Album*, 59 ALA. L. REV. 345 (2007) (arguing for abolition).

[97] *See* Samuelson, *supra* note 81, at 1525–27 (proposing to construe the derivatives right in light of the definitional examples); Christina Bohannan, *Taming the Derivative Works Right: A Modest Proposal for Reducing Overbreadth and Vagueness in Copyright*, 12 VAND. J. ENT. & TECH. L. 669, 696 (2010) (proposing that the definition of a derivative work "should be interpreted under the principle noscitur a sociis in light of the more specific examples that precede it").

[98] Gervais, *supra* note 81, at 805.

[99] *See* Naomi Abe Voegtli, *Rethinking Derivative Rights*, 63 BROOK. L. REV. 1213, 1267–68 (1997). *See also* Tim Wu, *Tolerated Use*, 31 COLUM. J. L. & ARTS 617, 630–31 (2008) (suggesting that the right of derivatives should not apply to complementary uses); Tyler T. Ochoa, *Copyright, Derivative Works and Fixation: Is Galoob a Mirage, or Does the Form(GEN) of the Alleged Derivative Work Matter?*, 20 SANTA CLARA COMPUT. & HIGH TECH. L.J. 991, 1044 (2004) (proposing to limit the right of derivatives to cases when one of copyright's other four exclusive rights are implicated).

[100] Voegtli, *supra* note 99, at 1264–65 (discussing a compulsory license scheme); Jed Rubenfeld, *The Freedom of Imagination: Copyright's*

B. A Right in Search of Justification

The debate about the scope of the derivatives right is grounded in a deeper difficulty: its precarious normative justification. It is often assumed that the right of derivatives is justified by the goal of allowing copyright owners to internalize the full social value of their works. As Professor Zechariah Chafee framed this argument long ago: "The essential principle is the author's right to control all the channels through which his work or any fragments of his work reach the market."[101] The full internalization idea seamlessly translates into the normative justification that the right of derivatives is desirable because it hitches the level of incentive to create as well as investment decisions by creators to the full social value of the work in all relevant markets.[102] The trouble with this argument is that there is no normative case for full internalization of the work's social value from any plausible normative perspective, whether based on incentive or fair compensation.[103] Since expressive works are strongly nonrival—many can use them without lessening the ability of others to do the same—there is no need for the governance function of property in allocating access to the work, and the only benefits in exclusion are with respect to the "dynamic" concerns of incentive to create and fair compensation.[104] The latter interests, however, do not require full internalization, which leaves no support for such a goal.

Once the erroneous assumption of the desirability of full internalization is removed, any other basis for the right of derivatives proves to be extremely weak. Traditional economic frameworks of copyright revolve around the goal of achieving a high

---

Constitutionality, 112 YALE L. J. 1, 55 (2002) (advocating for limiting the remedy to profit sharing); Orit Fischman Afori, *Flexible Remedies as a Means to Counteract Failures in Copyright Law*, 29 CARDOZO ARTS & ENT. L.J. 1, 32 (2011) (discussing derivative works in the context of the flexible use of remedies).

[101] Zechariah Chafee, *Reflections on the Law of Copyright: I*, 45 COLUM. L. REV. 503, 505 (1945).

[102] *See* Goldstein, *supra* note 91, at 227.

[103] *See* Oren Bracha & Talha Syed, *Beyond Efficiency: Consequence-Sensitive Theories of Copyright*, 29 BERK. TECH. L.J. 229, 295–96 (2014) (explaining why fair compensation to creators does not justify full internalization).

[104] Oren Bracha, *Give Us Back Our Tragedy: Nonrivalry in Intellectual Property Law and Policy*, 19 THEORETICAL INQUIRIES L. 633, 641–42 (2018).

"bang" of incentive-to-create benefit for a low "buck" by way of access cost.[105] The right of derivatives does poorly under this criterion. On the incentive benefit side, typically, commercially successful works will be able to recoup without the derivative-right-backed control of secondary markets, while additional profit streams for unsuccessful works will be minor (there is not much demand for action figures based on a resounding box-office failure).[106] At the same time, the right lays a heavy cost burden on secondary creation.[107]

Similarly, attempts to justify the derivatives right on product-differentiation grounds run into deep problems. The argument here is that the main function of the right of derivatives is to prevent rent dissipation: the wasteful race to capture demand by developing secondary works based on popular primary works.[108] Such races, as in the case of numerous computer games based on a successful movie, involve duplicative and socially wasteful costs of development incurred by many developers and only small social contributions through the variety provided by their competing products.[109] Seen through this prism, the role of the right of

---

[105] William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659, 1703–17 (1988) (explaining an incentive/loss ratio analysis for evaluating copyright entitlements); Louis Kaplow, *The Patent-Antitrust Intersection: A Reappraisal,* 97 HARV. L. REV. 1813, 1829–34 (1984) (developing an incentive/loss ratio analysis in the patent context).

[106] Stewart E. Sterk, *Rhetoric and Reality in Copyright Law*, 94 MICH. L. REV. 1197, 1216 (1996); Voeglti, *supra* note 99, at 1241–42; WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 109–10 (2003).

[107] *See* Voeglti, *supra* note 99, at 1269. Debate continues on the extent to which, in specific contexts, derivative rights have a better incentive/access ratio. *See* Samuelson, *supra* note 81, at 1527–33; Bambauer, *supra* note 96, at 381–82 (discussing the example of comic books where secondary markets' revenue may be substantial); Sterk, *supra* note 106, at 1216 (giving an example of an "extraordinarily high budget movie with the potential for sales of toys, t-shirts, and the like").

[108] *See* Michael Abramowicz, *A Theory of Copyright's Derivative Right and Related Doctrines*, 90 MINN. L. REV. 317, 358–59 (2005). *See also* Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265, 276 (1977).

[109] Abramowicz, *supra* note 108, at 358 (arguing that the concern "is that derivative works will be redundant with one another").

derivatives is to rationalize secondary works development by concentrating control and coordination power in the hands of the copyright owner in the primary work.[110] However, the argument runs out of steam when it is revealed that the solution simply shifts the problem: strong derivative rights to coordinate secondary works, increase the rents generated by the primary work, and consequently fuel the same kind of wasteful races on that level.[111] In the end, derivative rights is the wrong solution to the identified problem because it fails to address the source of the problem, namely property rights that enable the internalization of value and thereby generate rents.[112] Fixing a secondary development problem by strengthening property rights on the primary level "is the equivalent of fixing a hole in the floor by sawing a hole around it."[113] Rather, when the rent dissipation concern is dominant, the appropriate response is various ways to weaken the property right in the secondary creation.[114]

When one shifts their gaze to democratic or free-speech-based theories of copyright, the case for the derivative works right is further tarnished. Such theories go beyond market-based cost-benefit evaluation of the effects of copyright. They place a strong normative premium on a vibrant public sphere of expression consisting of a broad variety of sources and a meaningful diversity of viewpoints and alternatives.[115] In particular, this perspective values robust opportunities for the creation of speech that challenges broadly held views, beliefs, or tastes and is outside the mainstream, avant-garde, or is aimed at relatively marginal preferences.[116] The

---

[110] *Id.*

[111] *See* Bracha & Syed, *supra* note 50, at 1892–93. For the original critique in the patent context, see Donald G. McFetridge & Douglas A. Smith, *Patents, Prospects, and Economic Surplus: A Comment*, 23 J.L. & ECON. 197 (1980).

[112] *See* Bracha & Syed, *supra* note 50, at 1905; Bracha, *supra* note 104, at 662.

[113] Bracha, *supra* note 104, at 663.

[114] *See* Bracha & Syed, *supra* note 50, at 1910–18 (discussing alternative ways of reforming copyright to address rent dissipation problems).

[115] *See e.g.*, Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283 (1996).

[116] *See* Bracha & Syed, *supra* note 103 at 269–70 (discussing the concept of "heterodox works"); Netanel, *supra* note 115, at 40 (explaining the concept of "oppositional expression").

derivative works right does poorly in facilitating such an expressive environment. The reason: overall, the kind of heterodox works that are valued by democratic theory are likely to enjoy little of the right's benefit and disproportionally suffer its costs.[117] Being outside the mainstream and catering to marginal preferences makes the benefits of internalizing the value of secondary markets significantly smaller.[118] At the same time, such works are unlikely to generate enough profits to cover licensing costs while often needing to draw on culturally dominant materials to pursue their strategies of creative appropriation and user innovation.[119] Strong derivative rights are a sure recipe for structurally disadvantaging the kind of works that democratic theories value the most.

One could try to offer other grounds for the right of derivatives, perhaps based on a strong personal interest of creators in controlling the social meaning of their works, or in natural rights, labor-desert theories.[120] However, when examined closely, these justifications are not likely to fare much better than economic or democratic frameworks.[121] In short, the right of derivatives is a right in search of a justification. This precarious normative basis is closely related to the scope debate: the broader the asserted scope of the right of

---

[117] Bracha & Syed, *supra* note 103, at 272.

[118] *Id.* at 271.

[119] *Id.* at 271–72; Netanel, *supra* note 115, at 41 (arguing that oppositional speakers "are far less able to acquire copyright permissions"). On the role of creative appropriation in this context, see *id.*, at 159; Netanel, *supra* note 115, at 1221–26 (discussing "appropriative art"). On user innovation and its need to draw on preexisting cultural materials, see YOCHAI BENKLER, THE WEALTH OF NETWORKS 276 (2008); MADHAVI SUNDER, FROM GOODS TO A GOOD LIFE: INTELLECTUAL PROPERTY AND GLOBAL JUSTICE 105–14 (2012); Rosemary J. Coombe, *Objects of Property and Subjects of Politics: Intellectual Property Laws and Democratic Dialogue*, 69 TEX. L. REV. 1853, 1864 (1991); William W. Fisher, *The Implications for Law of User Innovation*, 94 MINN. L. REV. 1417, 1460 (2010).

[120] Bambauer, *supra* note 96, at 403–04.

[121] In brief: Personality-based justifications are likely to fail because it is impossible to adequately protect an inviolable, highly personal interest through alienable economic rights built around extraction of market value. Labor-desert arguments are likely to fail because there is no plausible argument that creators have an equitable claim for the full social value of their works.

derivatives, the harder it is to justify it and the greater are the concerns surrounding it.

C. Conceptual Difficulties

Related to the controversial scope of the right of derivatives and to its precarious normative basis is a host of conceptual and doctrinal puzzles of various specific aspects of the right.[122] It suffices to explain here the two most important puzzles that bear directly on the application of the right to the GenAI context.

The first conceptual difficulty is the obscure relationship between the right of reproduction and the right of derivatives. Where does the one end and the other start? No one seems to know, and few seem to care.[123] The reason for this obscurity is twofold. Looking at the question from one direction, the debate about the derivatives right's scope and its unclear normative grounding makes it hard to determine when exactly a particular use of a primary work enters the zone of derivatives. From the other side of the line, a tendency to expand the scope of the right of reproduction has further complicated the distinction. In modern copyright law, the scope of reproduction stretches beyond literal copying or even copying with trivial changes into a zone of substantial similarity.[124] And the penumbra of substantial similarity is generally recognized to be capacious and hard to pinpoint with accuracy.[125] Moreover, some courts have further diluted the substantial similarity requirement for

---

[122] Samuelson, *supra* note 81, at 1510.

[123] A notable exception is Professor Daniel Gervais, who analyzes in great detail the relationship between the right of reproduction and the right of derivatives. *See* Gervais, *supra* note 81, at 839–48.

[124] Arnstein v. Porter, 154 F.2d 464, 473 (2d Cir. 1946) ("The question . . . is whether defendant took from plaintiff's works so much of what is pleasing to the ears of lay listeners, who comprise the audience for whom such popular music is composed, that defendant wrongfully appropriated something which belongs to the plaintiff.").

[125] *See, e.g.*, Peter Pan Fabrics, Inc. v. Martin Weiner Corp., 274 F.2d 487, 489 (2d Cir. 1960) (noting that the test for infringement of a copyright is necessarily "vague" and determinations must be made "ad hoc"); Mark A. Lemley, *Our Bizarre System for Proving Copyright Infringement*, 57 J. COPYRIGHT SOC'Y U.S.A. 719, 720 (2010); Rebecca Tushnet, *Worth a Thousand Words: The Images of Copyright Law,* 125 HARV. L. REV. 683, 716–17 (2012).

infringement of the reproduction right, reducing it to an exception for de minimis takings, and have thereby further expanded that right's scope.[126]

With an unclear rationale for the right of derivatives and an ever-expanding frontier of reproduction, drawing the line becomes increasingly hard.[127] Some commentators struggle to clearly reconstruct the dividing line.[128] Others eschew the question by openly arguing that the right of derivatives is reducible to the reproduction right, or to reproduction and other rights.[129] Except in limited doctrinal contexts, where distinguishing the rights is inescapable, many courts appear to have little interest in drawing or maintaining a distinction.[130] This approach is mainly implicit in courts' tendency to not clearly differentiate the rights, or sometimes even to not clarify which right they analyze, perhaps under the assumption that the question is of little significance, given the mostly identical results of infringing either right.[131]

The second conceptual difficulty follows directly from the debates over the scope of the right. The difficulty is considerable obscurity on the issue of the outer boundaries of the right of derivatives. Courts and commentators often agree that the right of derivatives is not boundless, sometimes offering the example of a secondary work that is merely inspired by another as falling outside that scope.[132] However, in the wake of the freewheeling approach to

---

[126] Oren Bracha, *Not De Minimis: (Improper) Appropriation in. Copyright*, 68 AM. U. L. REV. 139, 164–69 (2018).

[127] *See* Oren Bracha & John. M. Golden, *Redundancy and Anti-Redundancy in Copyright*, 51 CONN. L. REV. 247, 297–98 (2019).

[128] *See id.*; Gervais, *supra* note 81, at 839–48.

[129] NIMMER & NIMMER, *supra* note 90, § 8.09[A][1]; Ochoa, *supra* note 99, at 1020.

[130] Bracha & Golden, *supra* note 127, at 298–99; NIMMER & NIMMER, *supra* note 90, § 8.09[A][1] (referring to "some exotic situations in which the adaptation right may take on substantive significance").

[131] Bracha & Golden, *supra* note 127, at 298.

[132] Warner Bros. Ent. Inc. v. RDR Books, 575 F. Supp. 2d 513, 538 (S.D.N.Y. 2008); NIMMER & NIMMER, *supra* note 90, § 8.09[A][1]; PATRY, *supra* note 88, § 12.12.

the scope of the right, it is often extremely hard to discern in a principled way when this outer limit is reached.[133]

The result of the two conceptual difficulties combined is something close to collapsing the two rights into one reproduction-derivative super-right. Within this one undifferentiated mass of a right to exclude, reproduction and derivatives blend into each other and expand outward to a considerable and nebulous extent.

V. THE RIGHT OF DERIVATIVES AND GENERATIVE AI INFRINGEMENT

Where the right of reproduction falters, the right of derivatives is called upon to establish broad GenAI copyright infringement liability, either upstream or downstream. This Part explains the different variants of broad derivatives infringement arguments. It then critiques these arguments by explaining how the logic of each leads to problematic or even absurd results that clash with copyright law's basic principles. Ultimately, these arguments run into deep trouble because in trying to evade the limitations of the right of reproduction, they rely on versions of the derivatives right that bring the inherent conceptual and normative problems of this right to an extreme. Put differently, to obtain broad GenAI liability, plaintiffs rely on the least plausible version of the right of derivatives.

A. Upstream

1. The Argument

With respect to upstream activities, whether training copies or model creation, the derivatives right argument is straightforward. It is that the very act of using a work for reproduction in training copies or the production of a model that is an infringing derivative use of the relevant works.[134] Although it is sometimes wrapped in elliptical

---

[133] Gervais, *supra* note 81, at 798.

[134] Kadrey v. Meta Platforms, Inc., No. 23-cv-03417-VC, 2023 WL 8039640, at *1 (N.D. Cal. Nov. 20, 2023) ("The plaintiffs allege that the 'LLaMA language models are themselves infringing derivative works' because the 'models cannot function without the expressive information extracted . . . .' "); Complaint at 29–30, N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Dec. 27,

or obscure language, the heart of the argument is simple. The purpose of the right of derivatives, the argument goes, is to allow the copyright owner to control and extract value from any market for any valuable use of the copyrighted work.[135] Reproduction in training copies and extraction of metainformation for model production are uses of the works in the training set. And these uses, being a crucial part of producing the immensely valuable GenAI system, have substantial value. The conclusion seems to inevitably follow: upstream activities related to copyrighted works are uses of the work that have market value potentially extractable by copyright owners; therefore, they are derivative uses within the confines of the right.

As for the remaining question of grounding the argument in the statutory text, the response may be that both training copies and models are works "based upon" copyrighted works.[136] Clearly, both training copies and models derive their informational content from the works on which they rely, by either literal reproduction or extraction of metainformation. How is the copyrighted work "recast, transformed, or adapted" by the derivative work? [137] Presumably, the transformation is by harnessing the information derived from the work to achieve a new purpose that is of considerable market value.

In short, the argument embodies a particularly refined form of the "if market value, then right" logic. Both the substantive assertion that any valuable use of a work is within the right and the matching definition of transforming a work reduce the inquiry to one of potential market value. It boils down to the assertion that if there is market value to any use of the work that is extractable by a right to exclude, then that use is captured by the right of derivatives.

2023) (alleging that "the GPT LLMs themselves have 'memorized' copies" of the work in the training set which are "unauthorized copies or derivative works").

[135] Goldstein, *supra* note 91, at 216.

[136] 17 U.S.C. § 101 (defining "derivative work"). *See* Lee et al., *supra* note 25, at 69 (arguing that a model "is more clearly a derivative work because it is 'based on' its training data").

[137] Lee et al., *supra* note 25, at 60.

2. The Argument Critiqued

The fatal flaw of the training copies and model as valuable uses argument is that it pulverizes copyright law's subject matter principles. The argument asserts that any valuable use of a copyrighted work is a derivative and, either obscures or directly rejects the relevance of the fact that the pertinent uses do not involve incorporation of protectable expression in a new expressive work. If this maneuver were to succeed, the result would be a blueprint for bypassing subject matter principles that limit copyright to the domain of expression. Broad upstream reproduction right arguments fail exactly because the uses they point at do not involve the taking of copyrightable subject matter.[138] If copyright owners could impose infringement by simply repackaging the same arguments under the right of derivatives, subject matter principles would be rendered meaningless.

The clash with subject matter principles is most obvious in the case of arguing that the model itself, being a valuable use that is based on works in the data set, is a derivative work. One court correctly dismissed such argument as "nonsensical."[139] It did so because it found there is no way of understanding the models themselves as a "recasting or adaptation" of the copyrighted works on which it was trained.[140] The more precise statement of this reason is that the model is not a recast or adapted version that incorporates any of the *expression* of the works in the training set. Rather, it is simply unprotected aggregate metainformation about the expressive works. Accordingly, asserting that such metainformation is a derivative is no more than a rhetorical maneuver designed to capture uncopyrightable subject matter.

The same reasoning applies to training copies. Under proper analysis, the reproduction argument fails with respect to non-expressive training copies because such reproduction does not involve any copyrightable subject matter.[141] As for the alternative ground under which training copies are exempted as fair use: at the

---

[138] *See supra* text accompanying notes 53–66.
[139] Kadrey v. Meta Platforms Inc., 2023 WL 8039640, at *1.
[140] *Id.*
[141] *See supra* text accompanying notes 53–62.

end, the heart of this reasoning too is subject matter principles, albeit repackaged into the fair use analysis factors.[142] Courts have consistently found non-expressive copies to be fair use exactly because neither the purpose of the reproduction nor their market effect have anything to do with appropriating a work's protected expression.[143] Again, simply reframing the infringement argument against training copies as being about a derivative use is no more than an evasion devoid of substance. The effect of such a move is to evade the heart of both the subject matter and fair use rejection of the reproduction right argument: namely, that the relevant reproduction does not appropriate the value of any copyrightable expression qua expression.

Right of derivatives arguments try to hop over these problems by a twofold strategy: pointing out in a loose way that models and training copies are derived or based on copyrighted works and placing the entire emphasis on the market value of this use of works. But none of this changes the fact that the use is not of copyrightable expression, or, to use the statutory language, does not involve transformation, recasting, or adaptation of expression.[144] In short, the upstream right of derivatives argument is the embodiment of the erroneous logic of full value internalization running amok.[145] Not only is there no reason in copyright policy for full internalization by the owner of the value of protected expression, but also in this case the argument is for owners internalizing value of non-expressive aspects of the work—an argument that, if accepted, would render subject matter principles null. In short, whatever one thinks about the proper scope of the right of derivatives, extending it to uses

---

[142] *See* Lemley & Casey, *supra* note 9, at 772–73, 775 (explaining that copying non-expressive subject matter, even when a verbatim non-expressive copy is made, does not involve the taking of copyrightable subject matter, conceding that it is reproduction nonetheless, and arguing that the taking is exempted as fair use); Samuelson, *supra* note 9 (manuscript at 71, 73) (explaining that "[t]he generative AI training process extracts information from millions or billions of works . . . to construct very different representations in models" and concluding that "the extractive purpose of training . . . favors generative AI fair use defenses").

[143] *See supra* note 39.

[144] 17 U.S.C. § 101 (defining "derivative work").

[145] *See supra* text accompanying notes 101–104.

based upon works but not involving the taking of any copyrightable subject matter must be a bridge too far.

B. Downstream

1. The Arguments

Downstream derivatives right arguments fall into two categories. One variant simply extends the logic of the upstream argument further down the line to encompass downstream generation. A second variant changes focus by training its sights on levels of expressive similarity that are insufficient to infringe the reproduction right but are claimed to be sufficient for purposes of the right of derivatives.

The first argument is that generated expressive output is a work which is, *in principle*, based on copyrighted works in the training set, one that recasts, transforms, or adapts these works.[146] The implicit claim is that generated output is derivative, with no need to show any specific similarity between such output and any copyrighted work.[147] Consider for example a generated image of two soccer players chasing a soccer ball.[148] Assume that the image does not bear concrete enough resemblance to a relevant work in the training set whose copyright is allegedly being infringed. However, the generated image does share certain patterns and general structural features with the copyrighted image as well as many other images in the training set. Indeed, it was produced by the GenAI

---

[146] *See* Tremblay v. OpenAI, Inc., No. 23-cv-03223-AMO, 2024 WL 557720, at *3 (N.D. Cal. Feb. 12, 2024) (arguing that "every output of the Open AI Language Models is an infringing derivative work . . . ."); Kadrey v. Meta Platforms, Inc., No. 23-cv-03417-VC, 2023 WL 8039640, at *1 (N.D. Cal. Nov. 20, 2023) (arguing that "every output of the LLaMA language models is an infringing derivative work"); Andersen v. Stability AI Ltd., No. 23-cv-00201-WHO, 2023 WL 7132064, at *1 (N.D. Cal. Oct. 30, 2023) (arguing that "every hybrid image is necessarily a derivative work" because "[e]very output image from the system is derived exclusively from the latent images").

[147] *See Andersen v. Stability AI Ltd.*, 2023 WL 7132064, at *1 (arguing infringement of the right of derivatives while admitting that "none of the Stable Diffusion output images provided in response to a particular Text Prompt is likely to be a close match for any specific image in the training data").

[148] *See* Complaint at 13, Getty Images (U.S.), Inc. v. Stability AI, No. 1:23-cv-00135-GBW (D. Del. Feb. 3, 2023).

system exactly by extracting these general patterns as metadata related to the expressive relations discernable in multiple works in the training set and then using them to generate a new concrete work.[149]

The right of derivatives argument is that a generated work is a derivative "based upon" a copyrighted work because these common patterns are derived *in part* from the copyrighted work. In a very loose sense, the claim is accurate: expressive GenAI is based on a process of extracting from concrete expression metainformation and then converting it back into concrete expression.[150] The metainformation represented in the model and pertaining to aggregate patterns and relations is derived from the works in the training set. Although it is impossible to trace any particular relation or pattern captured in the model to one specific work, in the aggregate, the model is derived from the works in the training set. Metaphorically, each work in the training set throws two cents into the big melting pot which produces the model. All the cents are then shuffled together, melted, and recast to produce the treasure which can be cashed out via generation. All works (together with the effort invested in the process) are responsible in the aggregate for the value of the metainformation in the model, but no specific feature of either the model or generated output can be traced specifically to one particular work.

The derivative right infringement argument says "no matter." Generated output contains structural elements that could be traced in a loose and aggregate way to a group of works in the training set. Such output is a valuable use whose expressive content is derived in, at least a small fraction, from each of the relevant concrete works in the set, therefore it is a derivative work.

The second argument is more specific. It does not encompass all generated works or even a group of works that incorporate some patterns that are conceivably traceable in small part to a specific copyrighted work in the training set. Instead, it applies to specific

---

[149] Sag, *supra* note 45, at 316 (explaining that "autoencoding is the process of abstracting latent features from the training data and then reconstructing those features").

[150] *See supra* Part II.A.

generated output that bears specific similarities to a particular work in the training set, albeit ones that are insufficient to satisfy the infringement test for purposes of the right of reproduction.[151] Reconsider the generated image of the soccer players. Compare it to a specific image in the training set and assume you can point out specific similarities between the two. Further assume that these similarities are insufficient to satisfy the infringement test under a right of reproduction claim because some similar elements are unprotectable, and any remaining commonalities of protected expression do not rise to the level of substantial similarity.[152] The argument is that these similarities are nonetheless sufficient to infringe the right of derivatives. The focus now is not so much on simply deriving valuable informational patterns from copyrighted works, but on claiming that the right of derivatives encompasses remote and diffused levels of similarity between two works. Put differently, the right of derivatives is claimed to be a right of reproduction on steroids.

2. The Arguments Critiqued

The first downstream derivatives right argument fails for the same reason as the upstream version. The argument simply extends the same logic of derivative uses untethered to appropriation of

[151] *See* Complaint at 33, N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Dec. 27, 2023) (arguing that "ChatGPT displays copies or derivatives of Times Works memorized by the underlying GPT models"). Of course, the claim of derivative output could be limited to such instances where there is substantial similarity between the output and the copyrighted works, but the point of adding the term "derivative" every time "copies" is mentioned seems to be exactly to expand the claim to cases of remote similarity. Anderson v. Stability AI, No. 23-cv-00201-WHO, 2023 WL 7132064, at *8 (N.D. Cal. Oct. 30, 2023) (expressing doubts that "copyright claims based on a derivative theory can survive absent 'substantial similarity' type allegations" but refusing to dismiss claims due to plaintiff's allegations that some output was substantially similar to plaintiffs' "style").

[152] Skidmore v. Led Zeppelin. 952 F.3d 1051, 1064 (9th Cir. 2020) (explaining the two parts of "unlawful appropriation" as: first, "because only substantial similarity in protectable expression may constitute actionable copying . . . 'it is essential to distinguish between the protected and unprotected material in a plaintiff's work;' " and second: establishing "similarity of expression from the standpoint of the ordinary reasonable observer, with no expert assistance").

specific expression to the generated output; and with this logic, it extends its fatal flaw. Attaching the claim of a derivative use that is based on a copyrighted work to generated output narrows the realm of application, but it does not change the untenable feature of the argument. Generated output is derived from works in the training set in the loose sense that it was produced by using patterns that are extracted in the aggregate from these works. But the output does not incorporate any concrete *expression* from any specific work. Extending the right of derivatives in this way to any valuable use of a work regardless of incorporation of specific expression risks once again running roughshod over subject matter principles.[153] If copyright subject matter restrictions are not to be nullified, the right of derivatives can only apply when specific expression that is traceable to a particular work, rather than unprotectable metainformation, is incorporated into the generated output.

The second downstream argument makes some progress by abandoning the attempt of trying to capture any valuable use and pointing at specific similarities between the copyrighted work and the derivative. However, by trying to base the claim of derivatives on similarities that do not satisfy the substantial similarity test, the claim risks hollowing out both scope and subject matter principles. Just as upstream arguments provided a blueprint for an end run around subject matter restrictions, this downstream argument purports to create a route for evading the copyright infringement test. This test is an important mechanism that embodies scope and subject matter principles.

To infringe the right of reproduction, it is insufficient for a work to copy from a copyrighted one.[154] In addition, the copying must

---

[153] *See* Tremblay v. OpenAI, Inc., No. 23-cv-03223-AMO, 2024 WL 557720, at *3 (N.D. Cal. Feb. 12, 2024) (finding an allegation that every output is a derivative work based on works in the training set "insufficient" due to plaintiff failing to argue "that any particular output is substantially similar—or similar at all" to their copyrighted works); Andersen v. Stability AI Ltd., No. 23-cv-00201-WHO, 2023 WL 7132064, at *8 (N.D. Cal. Oct. 30, 2023).

[154] Arnstein v. Porter, 154 F.2d 472, 472 (2d Cir. 1946) ("Assuming that adequate proof is made of copying, that is not enough; for there can be 'permissible copying.' ").

arise to the level of constituting improper appropriation.[155] This requires two distinct elements that at least some of the material copied is protectable expression, and that the protectable expression copied bears substantial similarity to the copyrighted work.[156] The infringement test is an important mechanism for safeguarding both scope and subject matter principles. Scope is the parameter that ensures copyright's balance between exclusion and access by setting the degree of similarity covered by the right.[157] The second element of the misappropriation requirement operationalizes this balance by mandating that only a taking that is substantially similar is infringing and allowing other copying.[158] Subject matter principles structure copyright law's balance on a more fundamental level by defining what kind of materials and activities are within the domain of copyright in the first place. The first element of the improper appropriation test bakes subject matter restrictions into the infringement analysis. Copyright protection is denied altogether to works that contain no copyrightable subject matter.[159] Filtering unprotectable elements as part of the infringement analysis operates on a complementary level: it ensures that copyright is not extended beyond its domain, even when a work contains protectable expressive elements, by insisting that it is only the taking of protected expression that can be the basis of copyright infringement.[160]

A right of derivatives argument that attempts to extend liability based on similarities that do not satisfy the infringement test circumvents the operation of that test and disrupts achieving its dual purpose. Under the logic of the argument, copying that is allowed by the infringement test under the right of reproduction would still be infringing under the right of derivatives, as long as any specific

---

[155] *Id.* at 468.

[156] Skidmore v. Led Zeppelin, 952 F.3d at 1064.

[157] *See* Pamela Samuelson, *A Fresh Look at Tests for Nonliteral Copyright Infringement*, 107 NW. U. L. REV. 1821, 1845–47 (2013).

[158] Bracha, *supra* note 126, at 175.

[159] 17 U.S.C. §102(b).

[160] *See* cases cited in *supra* note 62. *See also* Samuelson, *supra* note 157, at 1841–42 (discussing distinguishing protectable and unprotectable elements as part of the infringement analysis).

element is taken from the copyrighted work. An infringement test that could be so easily bypassed by a mere invocation of the right of derivatives would be of little significance and of no use in carrying out its function of regulating copyright's scope and domain coverage.

The point is not that there can be no overlap between the reproduction and derivatives rights. Some overlap or redundancy can be acceptable or even desirable in a well-designed copyright framework.[161] By contrast, it is not plausible for the right of derivatives to encompass all right of reproduction cases but extend protection to degrees of similarity and subject matter elements not covered by it. That would be an unlikely case of the right of derivatives devouring the right of reproduction and rendering inoperable its intricate infringement test and the functions it serves. In short, the argument for infringement of the right of derivatives by insufficiently similar generated output is the worst variant of fusing the reproduction and derivatives right into one super-right.[162] Such reasoning blurs the distinction between reproduction and derivatives and extends the scope of the joint right indefinitely.

VI. THE RIGHT OF DERIVATIVES TAMED

What is to be done about broad right of derivatives arguments, that stretch copyright beyond its limits and endanger basic principles of both subject matter and scope? This Part's argument is twofold. First, it explains that the case law already contains elements that could be used as local fixes to the GenAI right of derivatives challenge to copyright: existing doctrinal requirements that can be invoked to reject all versions of broad derivatives arguments. However, ultimately, GenAI right of derivatives arguments are a canary in a coal mine. The deep flaws of these arguments—in particular, the ways they clash with basic copyright principles—are warning signs about the deep problems of the right of derivatives in general. Dubious right of derivatives arguments in the GenAI context form a private case that reflects the general dangers of a broad and nebulous reading of the right of derivatives.

[161] Bracha & Golden, *supra* note 127, at 304.
[162] *See supra* text accompanying notes 122–32.

Consequently, this Part suggests a more general way of taming the right of derivatives: a reasoned elaboration of the right that avoids the most troubling dangers associated with it and mitigates its troublesome aspects.

A. Local Fixes

The right of derivatives case law contains elements that can be used to avoid the excess of broad arguments based on this right in the GenAI context. All arguments of this sort rely on one of two strategies. The first strategy is unmooring the concept of a derivative from any use of the expression of the copyrighted work. While the point is often made in an obfuscating way, its essence is simple: any valuable use of a copyrighted work is derivative, irrespective of the alleged derivative taking or incorporating specific protected expression from the primary work.[163] The effect is sweeping away almost any restriction on the concept of derivatives. The second strategy is somewhat more limited. It focuses on cases where expression was taken from the primary work but insists that a work can be an infringing derivative even if the expression taken bears only a remote level of similarity to the primary work—including similarity that is insufficient to satisfy copyright's general infringement test. As a result, infringement of the right of derivatives could be found even when the taken expression is of the kind that renders it unprotectable—or when it is protectable but not substantially similar to the primary work.[164]

Two strands in the case law provide an antidote to each of these strategies. With respect to the first strategy, the remedy is to insist that to be a derivative, a work must incorporate expression from the primary work. While some courts were willing to adopt, at least implicitly, a freewheeling conception of derivatives as any valuable use of the work, others have insisted that to be a derivative, a secondary work must incorporate expression from the primary work on which it is based.[165] The starting point for the latter position is

---

[163] *See supra* text accompanying notes 134–135.

[164] *See supra* text accompanying notes 151–152.

[165] A prime example of decisions that were willing to recognize as derivative a work that did not incorporate any expression of the primary work is this pair of cases: Worlds of Wonder Inc. v. Veritel Learning Sys. Inc., 658 F. Supp. 351 (N.D.

taking seriously the statutory definition. That definition does not stop with the general guidance that a derivative is "based upon" a preexisting work, but goes on to restrict the term to "any other form in which a work may be recast, transformed, or adapted."[166] The obvious meaning of this restricting clause, consistent with both copyright's general purpose and the legislative report, is that to recast, transform, or adapt a work means to incorporate some of its expression, while changing it in some way. This brings the definition into harmony with copyright law's basic subject matter principles that define its domain as expression.[167] And indeed, when discussing the derivative work right, the legislative report explains that "to constitute a violation . . . [a derivative] must incorporate a portion of the copyrighted work in some form."[168] A line of cases follow this understanding and refuse to recognize various uses of copyrighted

---

Tex. 1986); Worlds of Wonder Inc. v. Vector International Inc., 653 F. Supp. 135 (N.D. Ohio 1986) (ruling that audiotape cassettes with independently created stories that could be inserted into plaintiff's toy bear were infringing derivative works); *see also* Midway Mfg. Co. v. Artic Int'l, Inc., 704 F.2d 1009, 1013 (7th Cir. 1983) (finding an accelerator circuit board that when installed in plaintiff's arcade video game machine caused the game to play faster was an infringing derivative work); Micro Star v. FormGen Inc., 157 F.3d 1107, 1109–10 (9th Cir. 1998) (finding computer files containing specifications that when used with plaintiff's computer game caused it to generate new game levels infringing). For decisions that insist on a work incorporating expression to be a derivative see Ty, Inc. v. Publ'n Int'l, Ltd., 292 F.3d 512, 520 (7th Cir. 2002) (ruling that a collectors' guide of plaintiff's line of soft toys, as opposed to photographic representations of the toys, is not a derivative work because "guides don't recast, transform, or adapt the things to which they are guides"); Authors Guild v. Google, Inc., 804 F.3d 202, 225 (2d Cir. 2015) ("[D]erivative works over which the author of the original enjoys exclusive rights ordinarily are those that re-present the protected aspects of the original work, i.e., its expressive content."); Twin Peaks Prod., Inc. v. Publ'n Int'l, Ltd. 996 F.2d 1366, 1377 (2d Cir. 1993) (stating that an author "cannot preserve for itself the entire field of publishable works that wish to cash in on the" success of the work, but finding infringement due to taking of expression in television guide abridgement of copyright TV show episodes).

[166] *See* 17 U.S.C. § 101 (defining a "derivative work").

[167] *See id.* § 102(a) (providing that copyright subsists in "works of authorship"); *id.* § 102(b) (providing that copyright protection does not extend to various categories of non-expressive subject matter).

[168] *See* H.R. REP. NO. 941476, at 62 (1976). *See also* Samuelson, *supra* note 81, at 1526 (discussing the restriction in the legislative report and the limitations in the statutory definition).

works as derivatives when they do not involve incorporation of expression.[169] Thus for example, when the copyright owners in various Harry Potter works asserted that lexicon entries related to those works were infringing derivative works, one court ruled that a "work is not derivative . . . simply because it is 'based upon' the preexisting works" and insisted that to be a derivative, a work must be of the kind that is "representing the 'original work of authorship.' "[170]

An incorporation of expression requirement for derivatives, strikes at the root of broad GenAI derivative arguments, either upstream or downstream, that attempt to sweep in any valuable use of works, including extraction of unprotectable informational elements. Non-expressive physical training copies, "use" of metainformation extracted from a works to create a model, and further use of the model to generate new and different works, all do not involve incorporation of expression in the sense relevant to copyright law.[171] The first is a mere physical fact that involves no extension of access to the use value of the expression. The other two activities involve extracting non-expressive metainformation and using it to generate new expression. Consequently, none of these activities are the preparation of a derivative work.

As for the second strategy of extending the category of derivatives to takings of expressive material that bears only remote

---

[169] *See supra* note 163165. *See also* NIMMER & NIMMER, *supra* note 90, § 8.09 (observing that "[u]nless enough of the preexisting work is contained in the later work . . . by definition [it] is not a derivative work"); PATRY *supra* note 88, § 12.12 ("Without incorporation of a substantial amount of protectible material from a copyrighted work, a subsequent work cannot be considered a derivative work.").

[170] Warner Bros. Ent. Inc. v. RDR Books, 575 F. Supp. 2d 513, 538 (S.D.N.Y. 2008).

[171] *See* Kadrey v. Meta Platforms, Inc., No. 23-cv-03417-VC 2023 WL 8039640, at *1 (N.D. Cal. Nov. 20, 2023) (ruling that because a "derivative work is 'a work based upon one or more preexisting works' in any 'form in which a work may be recast, transformed, or adapted' . . . . There is no way to understand the LLaMA models themselves as a recasting or adaptation of any of the plaintiffs' books"). *See also* Samuelson, *supra* note 81, at 1545–46 (discussing computational extraction of metadata from texts and concluding that resultant sets of metadata "are not derivative works because they do not appropriate expression from protected works").

similarity to the primary work, the antidote is straightforward: insisting on subjecting the right of derivatives to the standard infringement test. Again, the case law contains inconsistent strands. Some courts were lax in applying the improper appropriation requirement of the infringement test to derivative works, or devised special diluted versions of the test for derivative rights cases.[172] Other decisions have adamantly insisted that the right of derivatives gets no special treatment: to infringe, the defendant's work must satisfy the infringement test, namely the incorporated expression must be protectable and must bear substantial similarity to the protected work.[173] For obvious reasons, in the GenAI context, firm insistence on applying the infringement test to the right of derivatives frustrates attempts to extend the right to remote levels of similarity, and prevents the evasion by such extension of subject matter and scope principles.[174]

---

[172] *See, e.g.*, Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc., 150 F.3d 132, 139 (2d Cir. 1998) (finding the standard infringement test unhelpful in cases when works are in "different genres and to a lesser extent because they are in different media" and replacing it with a qualitative/quantitative test); Mulcahy v. Cheetah Learning LLC, 386 F.3d 849, 853–54 (8th Cir. 2004) (ruling that a work "is an infringing derivative work if it copied or condensed the qualitative core of one marketable portion of" the copyrighted work and that it "may be found to be derivative even if it has 'a different total concept and feel from the original work' ").

[173] *See* Litchfield v. Spielberg, 736 F.2d 1352, 1357 (9th Cir. 1984) (characterizing plaintiff's argument that there is no need for substantial similarity to infringe the derivative right as claiming that this right "radically altered the protection afforded by the law of copyright" and dismissing it as "frivolous"); Well–Made Toy Mfg. Corp. v. Goffa Int'l Corp., 354 F.3d 112, 117 (2d Cir. 2003) (explaining that "even when one work is 'based upon' another" to be an infringing derivative such work must be substantially similar to the copyrighted work"); Kohus v. Mariol, 328 F.3d 848, 858 (6th Cir. 2003); *see also* Samuelson *supra* note 9, at 68 (observing that "courts have required a finding of substantial similarity in expression to infringe" the right of derivatives).

[174] *See* Andersen v. Stability AI, Ltd., No. 23-cv-00201-WHO, 2023 WL 7132064, at *8 (N.D. Cal. Oct. 30, 2023) (observing that "the alleged infringer's derivative work must still bear some similarity"); *Kadrey v. Meta Platforms, Inc*., 2023 WL 8039640, at *1 (ruling that "The plaintiffs are wrong to say that, because their books were duplicated in full as part of the LLaMA training process, they do not need to allege any similarity between LLaMA outputs and their books to maintain a claim based on derivative infringement"); Tremblay v. OpenAI, Inc.,

B. The Big Picture

Existing strands of case law can be marshaled to block broad GenAI derivative arguments. At the end, however, these local fixes are fingers in the proverbial dam that will keep threatening to break. Broad derivative arguments capitalize on deep ambiguities and difficulties associated with the right of derivatives in general. GenAI copyright disputes are simply a high-stakes context that motivates parties to push the derivative right to its limits and thereby expose its deep problems. These problems are traceable to two elements. The first is the obscurities associated with the right's justification, scope, and relationship with the right of reproduction.[175] The second is the erroneous but powerful idea that the right of derivatives is about internalizing the full social value of the work.[176] This notion generates pressure for promoting the least plausible versions of the right of derivatives. As long as these elements persist, confusion will abound, lawyers will keep trying to push implausibly broad versions of the right, and some courts might falter.

Moreover, in conjunction, the two local fixes to broad derivative arguments discussed above create a somewhat incoherent, or at least puzzling, conception of the right of derivatives. If the right of derivatives is limited to cases of incorporation of expression from the primary work and this taking must satisfy the improper appropriation requirement of the general infringement test, then how is the right of derivatives different from the right of reproduction?[177]

No. 23-cv-03223-AMO, 2024 WL 557720, at *3 (N.D. Cal. Feb. 12, 2024) (dismissing an argument that every output of a GenAI system is an infringing derivative work because the plaintiff failed "to explain what the outputs entail or allege that any particular output is substantially similar – or similar at all – to their books").

[175] *See supra* Part IV.

[176] *See supra* text accompanying notes 100–104.

[177] Bracketing the somewhat exotic group of cases where the preparation of the derivative work is done by way of altering a physical copy of the original and therefore involves no reproduction. *See* Mirage Editions, Inc. v. Albuquerque A.R.T. Co., 856 F.2d 1341 (9th Cir. 1988); Lee v. A.R.T. Co., 125 F.3d 580 (7th Cir. 1997). This limited group of cases hardly seems a plausible explanation for the significance of the right of derivatives.

It is probably this difficulty that led a central treatise writer to the conclusion that the right of derivatives is superfluous.[178]

A more principled solution would be to adopt a general proper conception of the right of derivatives, one that curbs the excesses of overbroad conceptions in a firm and systematic manner and also makes sense of the relationship between the rights of derivative and reproduction.[179] A good starting point for such a conception is the specific uses listed expressly in the statutory definition of the right.[180] As argued by Professor Pamela Samuelson, the "exemplary derivatives in the statutory definition" provide the basis for a coherent conception that is consistent with the statutory text, the legislative report, and the purpose of copyright in general.[181] This conception should be constructed not by mere analogies that form a disjointed list of features extracted from the various specified examples and are then sought after as traits of "family resemblance" in other potential uses. The preferable way is to use the specified cases as a foundation for a coherent and unified conception.

The key for building such a conception of derivatives is the first listed use: the translation.[182] What makes a translation an exemplary case of a unique category of secondary uses of works, one that is clearly distinct from both reproduction on the one hand, and any use of a work or even its expression, on the other? The answer is that a translation is a paradigmatic example of adaptation. As such, it has two defining features. On the one hand, an adaptation re-communicates or incorporates the entire expressive work, or a

---

[178] NIMMER & NIMMER, *supra* note 90, § 8.09[A][1].

[179] The Author describes here the general outline for a sound conception of the right of derivatives that Talha Syed and the Author elaborate on in greater detail elsewhere. *See* Talha Syed & Oren Bracha, Copyright Rebooted (unpublished manuscript) (on file with authors). This conception of the right as a right of adaptations is close to the suggestion of Professor Gervais of understanding the right as "asking whether what intrinsically makes the primary work original was taken, and then whether those elements were transformed." *See* Gervais, *supra* note 81, at 842.

[180] *See* Bohannan, *supra* note 97, at 696 (proposing a "narrower and more reasonable interpretation of the language" of the listed derivative uses to tame the right's breadth and vagueness).

[181] Samuelson, *supra* note 81, at 1511.

[182] 17 U.S.C. § 101 (defining a "derivative work").

primary part of it. We would not call a new work that does not re-communicate the novel *Gone with the Wind* a translation of that novel. This is the feature that distinguishes adaptations from the much larger set of all secondary uses. On the other hand, the adaptation recasts or transforms the work into a new expressive medium.[183] In the case of a translation, the changed medium is that of the language. The change of medium differentiates adaptation from reproduction. An adaptation is a close recasting of a work into a new medium. Translation is the paradigmatic case of adaptation and the first one listed in the statutory definition. However, all listed exemplars, motion picture versions or musical arrangements, for example, easily fit the mold, unless they are read in an extremely broad way.[184] The dual features of adaptation—recommunicating a work in a new medium—are also harmonious with the statutory text that defines the right beyond the listed exemplars as including: "any other form in which *a work* may be *recast, transformed, or adapted*."[185]

Some courts have read the right of derivatives exactly as a right of adaptation, including both features of rendering an *entire work* in a *new medium*.[186] The best example is *Penguin Random House LLC v. Colting*,[187] which captured how this conception tracks the specific exemplars in the statutory definition: "[l]ike a translation, dramatization, or motion picture adaptation (three categories explicitly delineated by Congress as derivative works, see 17 U.S.C.

---

[183] *See* Wu *supra* note 99, at 632 (proposing an approach to the derivatives right as "the right of adaptation between media").

[184] 17 U.S.C. § 101 (defining a "derivative work").

[185] *Id.*

[186] *See* Warner Bros. Ent. Inc. v. RDR Books, 575 F. Supp. 2d 513, 538 (S.D.N.Y. 2008) (holding that only works that are "recast, transformed, or adapted" into another medium, mode, or language while still representing the "original work of authorship" are derivative); Twin Peaks Prod., Inc. v. Publ'n Int'l, Ltd. 996 F.2d 1366, 1373 (2d Cir. 1993) (finding that a companion guide to a TV show is a derivative work because "[t]he Book contains a substantial amount of material from the teleplays, transformed from one medium into another"); Penguin Random House LLC v. Colting, 270 F.Supp.3d 736 (S.D.N.Y. 2017).

[187] Penguin Random House LLC v. Colting, 270 F.Supp.3d 736 (S.D.N.Y. 2017).

§ 101) . . . defendants' works basically *retell the story of plaintiffs' works* in *another medium*."[188]

Conceptualizing the right of derivatives as the right of adaptation results in a clear and coherent conception that significantly ameliorates the troublesome aspects of the right. It does so by addressing each of the three general difficulties associated with the right of derivatives.[189] First, the adaptation focus resolves the scope debate by setting a clear and well-circumscribed scope to the right. The right of adaptations is not a freewheeling entitlement that can be extended almost indefinitely to any use of a work. And the restriction is not via local, case-specific limitations or even an elusive list of analogous family resemblance features, but is rather based on a well-defined distinctive dual feature: the same work in a new medium. Second, the adaptation conception resolves the conceptual difficulties of the right about both its external boundaries and relationship with reproduction. Adaptation applies to a limited and well-defined subset of secondary uses and therefore no longer suffers from the problem of a nebulous and potentially ever-expandable outer boundary. At the same time, adaptation crisply defines the conceptual relationship to reproduction without eliminating the possibility of some overlap. Reproduction applies to identical or similar enough re-communication of a work embedded in a new physical object, while adaptation applies to re-communication of the work *in a new medium*.[190] Third, by ameliorating the scope and conceptual difficulties, adaptation significantly increases the normative plausibility of the right. No longer anchored in a full-internalization premise and limited to a significant but well-circumscribed additional zone of protection, the right of adaptation, while perhaps not foolproof, stands a much better chance of being justified either on incentive/access or fair compensation grounds.

As a bonus, a right of adaptations conception of the derivatives right also brings it into harmony with the relevant parts of the most

---

[188] *Id*. at 748 (emphasis added).

[189] *See supra* Part IV.

[190] *See* 17 U.S.C. § 106(1) (giving the owner the right "to reproduce the copyrighted work in copies"); *see also id.* § 101 (defining "Copies" as "material objects . . . in which a work is fixed").

important international agreement in the area of copyright: the Berne Convention.[191] The Convention requires giving authors a right of translation,[192] and a right of adaptation defined to include "adaptations, arrangements, and other alterations of works."[193] In other words, the Berne Convention's conception of the relevant right is not of an ever-extending sphere of derivatives, but expressly exactly that of well-defined adaptation with the translation being the paradigmatic member of the set.[194]

A right of adaptation conception addresses systematically the more general problems and dangers of the right of derivatives that the recent invocations of this right in GenAI copyright cases exposes. In the more specific context of GenAI infringement cases, the right of adaptation effectively curbs the excesses of broad right of derivatives arguments. GenAI derivatives arguments that try to extend the right to any use, even if unmoored from incorporation of expression, thereby endangering subject matter principles, are blocked. An adaptation requires recommunication of a work's expression, albeit in a different medium. Therefore, extraction of metainformation, mere physical aspects of this extraction, and use of the metainformation to generate new expression may be valuable uses of copyrighted works, but they are not adaptations of these works. Similarly, simply pointing at remote or diffused levels of similarities to copyrighted works in generated output—an argument that endangers subject matter and scope principles—is insufficient to establish the making of an adaptation. The latter requires recasting of the entire work or a primary part thereof, not merely some remote level of similarities in elements.

---

[191] Berne Convention for the Protection of Artistic and Literary Works, Sept. 9, 1886, *revised by* Paris Act on July 24, 1971 (amended July 24, 1979), 25 U.S.T. 1341, 828 U.N.T.S. 221.

[192] *Id.* art. 8.

[193] *Id.* art. 12.

[194] *See also id*. art. 2(3) (specifying that "[t]ranslations, adaptations, arrangements of music and other alterations of a literary or artistic work shall be protected as original works without prejudice to the copyright in the original work" and referring to this group of works as "[d]erivative works"); Gervais, *supra* note 81, at 820–22 (discussing the relevant rights under the Berne Convention).

Does this mean that the right of derivatives, properly conceived as the right of adaptation, has no work to do in the GenAI context? Not at all. Expressive GenAI systems already have uses that fall squarely within the right of adaptation. As these systems grow in power and range of application, the variety of such uses will increase. There are at least two patterns of GenAI generation that fit the mold of the adaptation right.

The first pattern is adaptation from prompt to output. In this pattern, the copyrighted expression is supplied by the prompt and therefore it is not necessary for the copyrighted work to be in the training set. The easiest example is that of translation. Prompting a GenAI system by using the text of a copyrighted poem that results in a generated translation of that poem is a clear case of adaptation. However, translation is only one example because the possible transformation of media that defines the adaptation is not limited to change of language. One can easily think of other examples, such as a copyrighted text prompt that results in a video rendering of the textual work (a motion picture version), a two-dimensional image prompt that results in the three-dimensional printing of a sculpture version (art reproduction), and so forth.

The second pattern is adaptation from training data to output via prompt. The media of adaptation in this pattern can be as diverse as in the first pattern. The difference is that the generated output is an adaptation of a copyrighted work in the training set, with the prompt only providing the guidance or trigger for the process. An easy example is prompting a system with "give me the lyrics of Bohemian Rhapsody in French," where the copyrighted English text was present in the training set and the generated output is a close French translation. The same pattern is extendable to other media.[195]

In short, a proper and properly restricted conception of the derivatives right as a right of adaptation has an important and distinctive role to play in the area of GenAI-produced expression, as it does elsewhere.

---

[195] *See, e.g.*, Gervais, *supra* note 11, at 1129 (discussing a scenario GenAI production of a 3D sculpture based on a copyrighted image in its training set).

VII. CONCLUSION

This Article argued that all variants of broad arguments of GenAI infringement of the right of derivatives stand on shaky ground and should be rejected. These arguments are devised as workarounds for bypassing the limitations of the reproduction right. In constructing broad right of derivatives arguments, plaintiffs are capitalizing on several dimensions of ambiguities associated with the right and are choosing the most capacious and loose version of conceptualizing the right on each. The result is an extremely implausible version of the derivatives right that clashes head on with copyright's general principles of subject matter and scope.

One set of arguments attempts to apply the right to any valuable use of copyrighted works as part of the GenAI production cycle, irrespective of incorporation of its expression in a new work. This makes the derivatives right a channel for evading subject matter restrictions. If all valuable uses of a work, including ones that only use unprotectable informational or physical aspects of works are infringing derivatives then little is left of subject matter principles that limit copyright's domain to expression. Another set of arguments asserts that any use of expression is derivative, irrespective of substantial similarity between the primary and secondary work. This logic causes the right of derivatives to swallow the right of reproduction. Thus construed, the derivatives right encompasses all cases of taking expression but disposes of any requirement of sufficient similarity to the copyrighted work. The result is endangering principles of both scope and subject matter.

The remedy for these ills of broad right of derivatives arguments is to firmly reject them. One route for achieving this result is existing strands of the case law that insist that to be a derivative, a work must incorporate expression from the primary work, and that to infringe, a derivative must meet the general substantial similarity test. A more principled alternative is to construe the right of derivatives as a right of adaptations, one that confers on owners a well-defined right to exclude secondary uses that incorporate their entire works while changing their medium.

Some argue that GenAI "Turns Copyright Upside Down."[196] The claim is that the way expressive output is produced by such systems challenges to the core fundamental assumptions and principles of copyright and will require their revision. Perhaps. Although it is just as likely that copyright principles, if applied correctly, can dispose of many of the seemingly challenging internal questions triggered by GenAI, while broader cultural policy concerns should be addressed outside of copyright.[197] The analysis of GenAI derivative right arguments reveals a different dynamic brought about by the technology. Here, the technological disruption creates disputes in a high stake-context that causes litigants to push existing legal doctrines to their limits. This reveals latent problems in the way these doctrines are understood and applied more generally in copyright law, which, in turn, behooves us to adjust the relevant doctrines to better conform with the fundamental principles of the field. In the case of the troublesome right of derivatives, this dynamic, set in motion by the GenAI challenge, brings home the point that the entitlement should be construed as the right of adaptations across the law of copyright.

[196] Mark A. Lemley, *How Generative AI Will Turn Copyright on its Head,* COLUM. SCI. & TECH. L. REV. (forthcoming 2024).

[197] *See* Bracha, *supra* note 2, at 23 (arguing that many broader cultural policy concerns that do not pertain to the public policy at the heart of copyright should be handled by other institutional mechanisms).