FAIR USE IN THE US REDUX:
REFORMED OR STILL DEFORMED?

Jane C. Ginsburg[*]

In 2019, Professor Ginsburg delivered the Distinguished Visitor in Intellectual Property Lecture at the Faculty of Law, National University of Singapore. Titled "Fair Use in the US: Transformed, Reformed, Deformed?", the lecture explored US caselaw applying the statutory fair use exception, highlighting its excesses and apparent rebalancing. Four and half years (and a pandemic) later, the Supreme Court has rendered decisions in two fair use cases (*Google v Oracle*; *Andy Warhol Foundation v Goldsmith*). Together, these controversies prompt inquiry whether the Supreme Court has redrawn the landscape of US fair use and copyright law, expanding fair use for commercial use of functional computer code, but narrowing it for at least some exploitations of "appropriation art." That inquiry extends to the fair use doctrine's potential to accommodate massive inputs of copyrighted works into databases to enable "machine learning" by artificial intelligence systems.

I. Introduction

Four and a half years ago, I gave a lecture at the Faculty of Law, National University of Singapore, taking the then-current pulse of the fair use defense in US copyright law. I suggested that after a period of nearly-invincible excess, the judge-made doctrine of "transformative use" might be tamed, as appellate courts expressed increasing skepticism over whether a given use was in fact "transformative," or ceased according determinative weight to the transformativeness of a given use.[1] In the intervening years, while lower federal courts continued to pull the fair use pendulum back toward equipoise, the Supreme Court's treatment of the doctrine has introduced new uncertainties, as the Court appeared to be enlarging the scope of the doctrine in its 2021 decision in *Google v Oracle*,[2] concerning the copying of functional computer code, but then confining it in its 2023 decision in *Andy Warhol Foundation v Goldsmith* ("*AWF*"),[3] concerning "appropriation art". Marking a second development since my pre-pandemic lecture, the proliferation of literary,

---

[1] Jane C. Ginsburg, "Fair Use in the United States: Transformed, Deformed, Reformed?" [2020] Sing JLS 265.
[2] *Google LLC v Oracle America, Inc* 141 US 1183 (2021) [*Google v Oracle*].
[3] *Andy Warhol Foundation for the Visual Arts v Goldsmith* 143 US 1258 (2023) [*Andy Warhol SC*].

artistic and musical outputs assisted by artificial intelligence ("AI") has roiled copy-right law, straining both the front end concept of authorship and the back end of copyright defenses, particularly fair use. I have written elsewhere about the front end.[4] In this article, I will confine the AI inquiry to the fair use status of the machine learning inputs: whether the massive ingestion of copyright works as "training data" does or should qualify as fair use.

The first part of this article will address the US Supreme Court's 2021 and 2023 decisions in *Google v Oracle* and *Andy Warhol Foundation v Goldsmith*. The second part will consider the application of fair use to the creation of databases to enable "machine learning".

II. RECENT SUPREME COURT CASELAW

A. *Functional Code: Google v Oracle*

In its 2021 decision in *Google v Oracle*,[5] the Supreme Court held that Google's copying of 11,500 lines of code from Sun Microsystems' Application Programming Interface ("API") in the development of Google's Android cell phone operating sys-tem was fair use. After failing to agree on a license from Oracle, Google copied the declaring code for 37 of Oracle's Java API packages and incorporated it into its own Android development platform.[6] Oracle's license terms required licensees to "share alike," that is, to make the code they develop from Oracle's code equally available to future downstream licensees, but Google preferred its code to be proprietary. Oracle also licensed its code without the "share alike" constraint but charged a higher fee that Google was unwilling to pay.[7] Google appropriated the code for the purpose of enticing software developers familiar with Oracle's API to use Google's Android platform.[8] Google hoped that if more developers started programming applications through Android, then its smartphones would be more attractive to consumers.[9] The Court assumed "for argument's sake" that the APIs were copyrightable,[10] but then devoted its fair use analysis to emphasizing its doubts about whether copyright

---

[4]   Jane C. Ginsburg & Luke A. Budiardjo, "Authors and Machines" (2019) 34(2) BTLJ 343.

[5]   *Supra* note 2.

[6]   *Ibid* at 1193.

[7]   See *Google v Oracle*, *ibid* at 1211, 1212, Thomas J, dissenting; *Oracle America, Inc v Google LLC*, 886 F 3d 1179 at 1187, 1188, rev'd 141 US 1183 (2021) [*Oracle v Google* (Fed Cit)].

[8]   *Supra* note 2 at 1194 (noting that "[w]ithout that copying, programmers would need to learn an entirely new system").

[9]   *Ibid* at 1190.

[10] *Ibid*. (Before turning to the merits of Google's fair use defense, the Court first considered Google's argument that the fair use question is one for juries–not judges–to decide (at 1199). In particular, Google asserted that fair use is a pure question of fact and that the Seventh Amendment, accordingly, required the Court to limit its scrutiny of the jury's determination to substantial evidence review, a highly def-erential standard (at 1199, 1200). Rejecting Google's argument, the Court read its precedents for the proposition that although application of the fair use defense requires the resolution of underlying factual issues, "the ultimate 'fair use' question primarily involves legal work"(at 1199, 1200). Accordingly, the Court agreed with the Federal Circuit that because fair use presents a mixed question of law and fact, the ultimate determination "is a legal question for judges to decide *de novo*"(at 1199). Assuming without

should cover the APIs. In effect, the fair use determination achieved the same result as ruling the APIs uncopyrightable, but attained that objective through the back end of a copyright exception rather than the front end of applying the idea/expression distinction to ascertain the scope of protectable expression: "[F]air use can play an important role in determining the lawful scope of a computer program copyright, such as the copyright at issue here."[11] The *Google v Oracle* court's constant references to the API's location far from the "core" of copyright prompts speculation that the fair use analysis masked a ruling on copyrightability for which a more forthright determination lacked a fifth justice.

Atypically for fair use analysis, but consistently with its backdoor assessment of copyrightability, the majority began its discussion with a lengthy analysis of the second fair use factor, "the nature of the copyrighted work," a factor that the last two and a half decades of fair use case law tended to recite and then ignore.[12] The Court noted that while Oracle's declaring code exhibited some creativity in its intuitive organization and easy-to-remember presentation, "its use is inherently bound together with uncopyrightable ideas (general task division and organization) and new creative expression (Android's implementing code)."[13] The Court, moreover, appeared especially concerned that the value of Oracle's declaring code derived substantially from the efforts of third-party developers to learn Oracle's system and create their own software products.[14] The majority's treatment of the second factor stressed that "the declaring code is, if copyrightable at all, further than are most computer programs (such as the implementing code) from the core of copyright"[15] (and, one may infer *a fortiori*, than more traditionally expressive works). The particularity of the nature of the copied code set it apart from other works and, some would contend, spawned a *sui generis* fair use analysis.[16]

The majority next turned to the first fair use factor, and inquired into the transformativeness of Google's copying of Sun's declaring code. The functional character of the declaring code informed the majority's analysis of the purpose and character of Google's use:

---

deciding that Oracle's API packages were copyrightable (see at 1197), the Court therefore proceeded to determine the fair use question for itself.)

[11] *Ibid* at 1198.

[12] The second factor weighed most heavily when the plaintiff's work was unpublished (see *Harper & Row Publishers, Inc v Nation Enterprises* 471 US 539 (1985) at 563, 564 [*Harper & Row*]), but a subsequent amendment to Copyright Law of the United States 17 USC (US) § 107 [Copyright Law of US] in response to lower court decisions overemphasizing works' unpublished nature clarified that a work's unpublished status is not dispositive.

[13] *Supra* note 2.

[14] See *Google v Oracle*, *ibid*. User investment in learning the program largely motivated the First Circuit's determination in *Lotus Development Corp v Borland International, Inc* 49 F 3d 807 (1st Cir, 1995) [*Lotus Dev Corp*] at 815–819 that the Lotus spreadsheet's menu command sequence was not copyrightable. The Supreme Court granted *certiorari* in that case, but, with one Justice recused, split 4–4, thus failing to render a precedential decision in the case (See *Lotus Development Corp v Borland International, Inc* 516 US 233 (1996)).

[15] *Supra* note 2 at 1202.

[16] See *Google v Oracle*, *ibid* at 1218 n 11, Thomas J, dissenting: "Because the majority's reasoning would undermine copyright protection for so many products long understood to be protected, I understand the majority's holding as a good-for-declaring-code-only precedent."

Here Google's use of the Sun Java API seeks to create new products. It seeks to expand the use and usefulness of Android-based smartphones. Its new product offers programmers a highly creative and innovative tool for a smartphone environment. To the extent that Google used parts of the Sun Java API to create a new platform that could be readily used by programmers, its use was consistent with that creative 'progress' that is the basic constitutional objective of copyright itself.[17]

Google's use of the code was "transformative" because "reimplementing an interface can further the development of computer programs."[18] Google copied Oracle's APIs only to the extent necessary to afford third-party programmers a familiar development environment on Google's platform.[19] To that end, it had repurposed for use in smartphones an API originally developed for use with legacy computers.[20] Altogether, the Court viewed Google's reimplementation of the API as the sort of reasonable use of a functional industry standard that is necessary to foster innovation and competition.[21] In assessing the transformativeness of Google's use of Oracle's declaring code, the Court kept evoking that code's diminished entitlement to copyright in the first place.With respect to the third fair use factor, the Court found that the amount and substantiality of the appropriated declaring code also favored fair use.[22] The Court declined to view "in isolation" the 11,500 lines of declaring code that Google copied, instead underscoring the 2.86 million lines of API code that Google *did not* copy.[23] The 11,500 lines "should be viewed … as one part of the considerably greater whole"[24] because of the peculiar nature of the copyrighted work, and the concomitant, transformative, purpose of the defendant's work. In other words:[25]

Google copied those lines not because of their creativity, their beauty, or even (in a sense) because of their purpose. It copied them because programmers had already learned to work with the Sun Java API's system, and it would have been difficult, perhaps prohibitively so, to attract programmers to build its Android smartphone system without them.

---

17 *Ibid* at 1203.
18 *Ibid*.
19 See *Google v Oracle*, *ibid*.
20 See *Google v Oracle*, *ibid*.
21 See *Google v Oracle*, *ibid* at 1203, 1204 (collecting briefs by *Amici Curiae* in support of Google).
22 *Ibid* at 1206.
23 See *Google v Oracle*, *ibid* at 1204, 1205. This approach is in some tension with traditional copyright doctrine. See *eg*, *Sheldon v Metro-Goldwyn Pictures Corp* 81 F 2d 49 (2nd Cir, 1936) at 56: "[N]o plagiarist can excuse the wrong by showing how much of his work he did not pirate." See also *Harper & Row*, *supra* note 12 at 600 (observing that although the defendants copied only 400 words of a many-thousand word book, these constituted "the heart of the book"). Copyright Law of US, *supra* note 12, § 107(3), however, directs inquiry into the "amount and substantiality of the portion used *in relation to the copyrighted work as a whole*" [emphasis added]. See also *Fioranelli v CBS Broadcasting Inc*, 551 F Supp 3d 199 (S.D.N.Y. Dist Ct, 2021) at 226 (declining to follow defendant's "purely mathematical approach" to the amount and substantiality, and duration, of copying from plaintiff's photographs into defendant's documentary films).
24 *Supra* note 2 at 1205.
25 *Ibid* at 1203.

Once again, the dubious copyrightability of functional declaring code permeated the Court's analysis not only of the first and second fair use factors, but also of the substantiality of the appropriation (factor three), and, as we shall next see, the impact of the use on the market for the copyrighted work (factor four).

Turning to the "market effects" of Google's copying, the Court began by explaining that while lost revenue provides one important measure of "market effects," courts must also consider "the source of the loss."[26] Citing its decision in *Campbell v Acuff-Rose*,[27] the Court stressed that certain types of market losses, such as those resulting from a "lethal parody" that "kill[s] demand in a work," have never been "cognizable under the Copyright Act." Additionally, the Court emphasized that any lost revenue must be weighed against any "public benefits the copying will likely produce."[28] In other words, revenue lost for non-copyright reasons (such as devastating criticism) does not count, and even where the copying can give rise to cognizable harm, the significance of the harm may wax or wane with the substantiality of a competing public interest.

The Court's inquiry into the impact of the copying followed that two-pronged framework. First, the Court called into question a causal relationship between Google's copying and harm to Oracle. Explaining that Oracle "was poorly positioned to succeed in the mobile phone market," the Court noted that the jury could reasonably have concluded that the company "would not have been able to enter those markets successfully whether Google did, or did not, copy a part of its API."[29] Similarly, the Court asserted that the evidence adduced at trial indicated "that Android was not a market substitute for Java's software" because "devices using Google's Android platform were different in kind from those that licensed [Oracle's] technology."[30] Specifically, the Court observed that while devices running Oracle's licensed technology tended to be "simpler products," more advanced devices tended to build on Google's Android platform.[31] Consequently, "[Oracle's] mobile phone business was declining, while the market increasingly demanded a new form of smartphone technology that [Oracle] was never able to offer."[32] Lastly, the Court seized on Oracle's apparent belief that it would derive "a benefit from the broader use of the Java programming language in a new platform like Android, as it would further expand the network of Java-trained programmers."[33] The Court also dismissed Oracle's claims of lost licensing revenue on the ground that the license it offered Google covered more than the code Google actually copied.[34] Ultimately, "neither Sun's effort to obtain a license nor Oracle's conflicting evidence can overcome evidence indicating that, at a minimum, it would have been difficult for Sun

---

[26] *Ibid* at 1206.

[27] *Campbell v Acuff-Rose Music, Inc* 510 US 569 (1994) at 591, 592 [*Campbell*]).

[28] *Google v Oracle, supra* note 2 at 1206.

[29] *Ibid*.

[30] *Ibid* at 1206, 1207.

[31] *Ibid* at 1207: "[R]ather than just repurposing [Oracle's] code from larger computers to smaller computers, Google's Android platform was part of a distinct (and more advanced) market than Java software." (internal quotation marks and citation omitted).

[32] *Ibid*.

[33] *Ibid*.

[34] *Ibid*.

to enter the smartphone market, even had Google not used portions of the Sun Java API."[35] Moreover, while Google may have made a great deal of money from its unlicensed use of the API, the Court again stressed that the API became "valuable [to Google] … because users, including programmers, are just used to it. They have already learned how to work with it… We have no reason to believe that the Copyright Act seeks to protect third parties' investment in learning how to operate a created work."[36]

With respect to the second prong, addressing the public interest at stake in the case, the Court opined that "given programmers' investment in learning the Sun Java API, to allow enforcement of Oracle's copyright here would risk harm to the public."[37] In other words, because Oracle's API had become an industry standard to which software developers had grown accustomed, coding an alternative system would likely impose great cost and difficulty.[38] For that reason, the Court feared that permitting Oracle a monopoly on its largely functional API might well stifle "creative improvements, new applications, and new uses developed by users who have learned to work with that interface."[39] In that case, a finding against fair use "would interfere with, not further, copyright's basic creativity objectives."[40]

This characterization of the public interest, however, arguably fails to take due account of the impact of Google's use on Oracle's "share alike" model. Recall that under this licensing scheme, Google would have been free to use Oracle's API packages so long as Google made any software incorporating Oracle's code compatible with other Java programs.[41] This model thus promotes device and software interoperability, broadening the public's access to useful technology. However, because Google wanted its platform to remain proprietary, it refused to accept this license and instead opted to appropriate Oracle's code anyway, thereby potentially limiting the availability of its software to the public for off-platform uses. As a result, it is not entirely clear that it was Oracle–and not Google–who threatened the public interest in this case. But this wrinkle notwithstanding, the Court, weighing the relevant considerations, concluded that the market effects factor also favored fair use.[42] As a result, with all four statutory factors favoring fair use, the Court found in favor of Google.[43]

---

[35] *Ibid*.

[36] *Ibid* at 1207, 1208 citing *Lotus Dev Corp*, *supra* note 14 at 821 (Boudin J, concurring).

[37] *Ibid* at 1208.

[38] See *Google v Oracle*, *ibid*.

[39] *Ibid*.

[40] *Ibid*.

[41] See *Google v Oracle*, *ibid* at 1212 (Thomas J, dissenting); *Oracle v Google* (Fed Cit), *supra* note 7 at 1350: "The point of contention between the parties was Google's refusal to make the implementation of its programs compatible with the Java virtual machine or interoperable with other Java programs. Because Sun/Oracle found that position to be anathema to the 'write once, run anywhere' philosophy, it did not grant Google a license to use the Java API packages."

[42] *Supra* note 2 at 1208.

[43] *Ibid* at 1208, 1209.

1. *Google's potential impact on fair use case law in general*

In many respects, *Google v Oracle* reprises in fair use guise the debates the Court failed to resolve 25 years earlier, in *Lotus v Borland*,[44] concerning the copyrightability of certain functional aspects of a user interface. In that case, the market-dominance of the Lotus spreadsheet led to its menu commands becoming the industry standard. The First Circuit ruled the commands an uncopyrightable "method of operation."[45] Judge Boudin's concurrence, thrice cited in Justice Breyer's majority opinion in *Google v Oracle*, explicitly justified the outcome on the ground that Lotus had become a *de facto* standard.[46] Judge Boudin also suggested that devising an exception for Borland's value-added copying might be preferable to holding the menu commands uncopyrightable, but, uncertain that Borland's commercial purpose would qualify as fair use, agreed with the determination to deny copyright protection altogether.[47]

The specificity of the context of Google–functional subject matter of borderline copyrightability, and its status as an industry standard – raises questions about the likely impact of the Supreme Court's decision on the general development of the fair use doctrine. As we have seen, the Court's decision in Google's favor on the second statutory fair use factor–the nature of the copyrighted work–permeated, if not controlled, its analysis of the other three. One should keep in mind the centrality of the second fair use factor to Google's analysis when contemplating that decision's potential impact on other fair use controversies.

For example, in its discussion of the first factor, the Court endorsed Google's use of the APIs "to create new products."[48] Were the Court's statements taken out of context, so that *verbatim* copying "to create new products" were deemed "transformative" in general, it would be difficult to imagine what kind of copying, short of outright piracy of the entire work, would not be transformative. Similarly, the Court's discounting of Oracle's cognizable harm on the ground that it was unlikely itself to develop a cellphone platform, would be very problematic were it extended to works "closer to the core of copyright." For example, at least until now, a film producer unwilling to purchase film rights from a novelist, and who makes the movie nonetheless, would not likely succeed in contending that an author who is unable to make the motion picture herself incurs no cognizable economic harm under the fourth fair use factor.[49]

One might counter that the Google majority's discussion of transformative use cites examples drawn from more traditional works of authorship. "[W]e have used the word 'transformative' to describe a copying use that adds something new and important. An 'artistic painting' might, for example, fall within the scope of fair use even though it precisely replicates a copyrighted 'advertising logo to make a

---

[44] *Lotus Dev Corp*, *supra* note 14.

[45] *Ibid* at 815–819.

[46] See *Lotus Dev Corp*, *ibid* at 821, 822 (Boudin J, concurring).

[47] See *Lotus Dev Corp*, *ibid*.

[48] *Supra* note 2 at 1203.

[49] See *supra* note 2 at 1217 (Thomas J, dissenting) (noting that "[a] book author need not be able to personally convert a book into a film so long as he can license someone else to do so").

comment about consumerism.'"[50] Arguably, this boiler-plate recitation endorses the kind of copying in which Warhol engaged in *AWF*,[51] which the Court decided two years later. But, with respect to Lynn Goldsmith's photograph of the performer Prince, to which Warhol applied his "flattening" silk screening technique, unlike the Campbell's Soup logo to which the *Google v Oracle* majority implicitly referred, Warhol was not "mak[ing] a comment" about what Goldsmith's work stood for. His technique may have "added something new and important," but were every "copying use that adds something new and important" to "fall within the scope of fair use," then virtually no copying, beyond the most blatant piracy, would fall outside the fair use shelter. As we will see, the Supreme Court in *AWF* declined to take its aside in Google out of context to create a broad fair use carve out for works that "add[] something new and important."[52]

One could espouse a principled position that "new and important" additions to copyrighted works should not infringe;[53] that the scope of copyright protection should be limited to *verbatim*, piratical copying. Such a position, however, is not the one Congress chose when it specified exclusive rights over derivative works, and when it directed courts to take into account not only the purpose and character of a defendant's use, but also the amount and substantiality of the use, and the effect of that use upon the potential market for or value of the copyrighted work.[54] Many if not most derivative works "add something new and important" to the works they copy and adapt; if that were all that was required to render the use "fair," then the use "if it should become widespread, it would adversely affect the potential market for the copyrighted work"[55] by usurping derivative works markets. One may therefore expect that works closer to the creative "core" of copyright will significantly stunt *Google v Oracle's* reach. On the other hand, the decision leaves open the possibility that Google's approach will bear more heavily when more factual or functional works are at issue. The Supreme Court's subsequent decision in *AWF*, confirms the limited reach of *Google v Oracle* to creative works at the "core" of copyright.[56]

---

[50] *Ibid* at 1203 (citing *Campbell*, *supra* note 27 at 579 and David Nimmer, *Nimmer on Copyright* (Matthew Bender Elite Products, 2019) at §13.05[A][1][b]).

[51] See Petition for Panel Rehearing and Rehearing En Banc at 2, 11, 12 in *Andy Warhol Foundation for Visual Arts v Goldsmith* No 19-2420-cv, 11 F 4th 26 (2nd Cir, 2021): "Indeed, *Google* described—as a paradigm example of transformative use—a Warhol-like work of art that is materially indistinguishable from the works at issue here." [*Andy Warhol II*].

[52] *Andy Warhol SC*, *supra* note 3 at 1299.

[53] Although such a standard would probably spur contention over what additions are "new and important," thus inviting assessments of the merits of the defendant's work.

[54] Copyright Law of US, *supra* note 12, §§ 106(2), 107(1), 107(3), 107(4).

[55] *Sony Corporation of America v Universal City Studios* 464 US 417 (1984) at 451.

[56] *Supra* note 3 at 1312 (The *AWF* majority confined *Google v Oracle* to fair use analysis of functional works. The majority's brief substantive discussion of *Google v Oracle*, in a footnote, reminded that "*Google* stressed that '[t]he fact that computer programs are primarily functional makes it difficult to apply traditional copyright concepts in that technological world.'" (at 1312 n 8)).
Lower courts have also limited *Google v Oracle* to apply to software, stating that its distinction between functional and expressive works favors fair use in the former instance, but not in the latter. Compare *Apple Inc v Corellium, Inc* No 21-12835, 2023 WL 3295671 (11th Cir, 2023) at *9 (software) with *Campbell v Gannett Co, Inc* (No 4:21-00557-CV-RK, 2023 WL 5250959 (W.D. Mo. Dist Ct, 2023) at 6 (photograph).

B. *Appropriation Art: Andy Warhol Foundation v Goldsmith*

AWF,[57] decided by the U.S. Supreme Court on May 18, 2023, pitted claims to free artistic use of source works against the ability of those works' creators (predominantly photographers) to exploit markets for works based on their creations. The US Copyright Act's provision for the making of derivative works gives the author (or successor in title) exclusive rights over "any… form in which the work may be recast, transformed or adapted."[58] Like all the Act's exclusive rights, the derivative work right is "subject to" the Act's exceptions and limitations, notably fair use. An unauthorized derivative work, such as an adaptation, musical arrangement, or editorial revisions constituting an original work of authorship, may be substantially similar to the underlying work and thus *prima facie* infringe its copyright, yet ultimately not give rise to liability for copyright infringement if it is a fair use. When the contested use incorporated copied content into a new work, caselaw in the lower federal courts after the Supreme Court's 1994 decision in *Campbell v Acuff-Rose*,[59] involving a musical parody of the Roy Orbison song "Pretty Woman," concentrated on the first factor. Courts inquired whether the defendant's use was "transformative," "add[ing] something new, with a further purpose, or different character, altering the first with new expression, meaning or message."[60] For some lower courts, "new meaning or message" became a mantra, a "get out jail free" card that increasingly risked carving derivative works out of the scope of exclusive rights.[61] But that result goes too far: that the statute "subject[s]" the derivative work right to the fair use exception, should not mean that all derivative works are automatically fair use. The Supreme Court has restored its original meaning of "transformative use," as a consideration to be weighed against other elements, notably the commercial character of the defendant's use. By emphasizing the purpose and character of the defendant's *use*, rather than focusing on whether the purpose and character of the defendant's *work* differ from the plaintiff's, the Supreme Court oriented the inquiry around the extent to which the mode of exploiting the defendant's work substitutes for one of the plaintiff's exploitations. The exploitation need not be the primary mode; the inquiry concerns secondary as well as primary markets.

The decision not only steers clear of invoking the artistic merits of the defendant's work; it also renews attention to the importance of copyright to the creators of the source works. The majority focused on the prospects of the first artist to make a living ("even against famous artists"),[62] while the dissent trained on the genius of

---

[57] *Supra* note 3.

[58] *Supra* note 12, §§ 106(2), 101 (definition of "derivative work").

[59] *Supra* note 27.

[60] *Ibid* at 579.

[61] *Dr. Seuss Enterprises v ComicMix LLC* 983 F 3d 443 (9th Cir, 2021) at 453 [*Dr Seuss*].

[62] *Supra* note 3 at 1287. See also *Dr Seuss*, *ibid* at 1278. "[Artist's reference] licenses, for photographs or derivatives of them, are how photographers like Goldsmith make a living. They provide an economic incentive to create original works, which is the goal of copyright." And *ibid* at 1286: "It will not impoverish our world to require AWF to pay Goldsmith a fraction of the proceeds from its reuse of her copyrighted work. Recall, payments like these are incentives for artists to create original works in the first place."

the second artist, and a long tradition of artistic borrowing from prior works.[63] The controversy began as a declaratory judgment action brought by the Andy Warhol Foundation ("AWF") against photographer Lynn Goldsmith, who had created a portrait of entertainer Prince in 1981. Goldsmith, in a "one time, one use" agreement, licensed the photograph in 1984 for $400 "to Vanity Fair magazine for use as an artist reference."[64] Vanity Fair commissioned Andy Warhol to create an illustration based on the photograph and it published Warhol's illustration to accompany an article about Prince in the November 1984 issue of the magazine.[65] The illustration published in Vanity Fair was one of a series of 16 silkscreen paintings, prints and drawings Warhol created based on Goldsmith's photograph, but that he did not sell or otherwise exploit during his lifetime.[66] After Prince's death in 2016, Vanity Fair obtained a license from AWF to republish one of the Warhol images (a different one than the magazine had printed in 1984) on the cover of a special issue of the magazine devoted to the performer. Vanity Fair did not obtain a license from Goldsmith, nor did its special issue credit her source photograph, although the November 1984 issue had included a source credit.[67]

Photo: Lynn Goldsmith[68]

---

[63] Although many of the dissent's references exemplify copying of ideas or genres (*eg*, reclining nudes) rather than the takings of specific expression which could constitute *prima facie* infringement to which the fair use defense might apply. The dissent broadly endorsed finding copying "transformative," if the defendant made creative use of the copied content, notably by adding new expression.

[64] *Andy Warhol Foundation for the Visual Arts, Inc v Goldsmith* 992 F 3d 99 (2nd Cir, 2021) [*Andy Warhol I*], opinion withdrawn and superseded on rehearing; *Andy Warhol II*, *supra* note 51.

[65] *Ibid*.

[66] *Ibid*.

[67] *Ibid*.

[68] *Andy Warhol SC*, *supra* note 3 at 1267.

Purple Prince: published in Vanity Fair in 1984[69]

Orange Prince: published in Vanity Fair in 2016[70]

The district court upheld AWF's fair use defense, finding the Warhol silkscreens transformative ("each Prince Series work is immediately recognizable as a 'Warhol' rather than as a photograph of Prince") and unlikely to supplant the market for Goldsmith's photograph ("[i]t is plain that the markets for a Warhol and for a

---

[69]   *Andy Warhol SC, ibid* at 1269.
[70]   *Ibid*.

Goldsmith fine-art or other type of print are different").[71] The court also gave short shrift to Goldsmith's contention that AWF's unlicensed use competed with her ability to license her photograph: "this does not suggest that a magazine or record company would license a transformative Warhol work in lieu of a realistic Goldsmith photograph."[72]

The Second Circuit Court of Appeals reversed.[73] Addressing the first fair use factor, the Second Circuit chided the district court for applying a bright line rule "that any secondary work that adds a new aesthetic or new expression to its source material is necessarily transformative."[74] Equally importantly, the court addressed the relationship between transformative fair use and the author's exclusive right to make or to authorize the making of derivative works. The court did not exclude all possibility that reworking a single prior work might yet be "transformative," but "the secondary work's transformative purpose and character must, at a bare minimum, comprise something more than the imposition of another artist's style on the primary work such that the secondary work remains both recognizably deriving from, and retaining the essential elements of, its source material."[75] The Second Circuit underscored:

> Finally, we feel compelled to clarify that it is entirely irrelevant to this analysis that "each Prince Series work is immediately recognizable as a 'Warhol.'" Entertaining that logic would inevitably create a celebrity-plagiarist privilege; the more established the artist and the more distinct that artist's style, the greater leeway that artist would have to pilfer the creative labors of others. But the law draws no such distinctions; whether the Prince Series images exhibit the style and characteristics typical of Warhol's work (which they do) does not bear on whether they qualify as fair use under the Copyright Act.[76]

The Second Circuit also observed that Warhol's use was "commercial in nature, but … produce[d] an artistic value that serves the greater public interest.… Nevertheless, just as we cannot hold that the Prince Series is transformative as a matter of law, neither can we conclude that Warhol and AWF are entitled to monetize it without paying Goldsmith the 'customary price' for the rights to her work."[77]

The Second Circuit also found the remaining fair use factors favored Goldsmith. Her work was creative (factor 2); Warhol copied the identifiable essence of Goldsmith's photograph without establishing a need to take *Goldsmith's* representation (as opposed to any photographic representation) of the artist Prince (factor 3); Warhol usurped the established market for licensing photographs as "artists' references" for magazine publication (factor 4).

---

[71] *Andy Warhol Found. for the Visual Arts v Goldsmith* 382 F Supp 3d 312 (S.D.N.Y. Dist Ct, 2019) 326 at 330.

[72] *Ibid* at 330, 331.

[73] *Andy Warhol II*, *supra* note 51.

[74] *Ibid* at 38, 39.

[75] *Ibid* at 42.

[76] *Ibid* (citation omitted).

[77] *Ibid* at 44.

The Supreme Court granted *certiorari*, but only as to the first factor, which it examined from the perspective of AWF's licensing of the work for publication in a magazine tribute to Prince.[78] "On that narrow issue, *and limited to the challenged use*, the Court agrees with the Second Circuit: The first factor favors Goldsmith, not AWF."[79] The Court rejected AWF's contention that "the Prince Series works are 'transformative,' and that the first factor therefore weighs in its favor, because the works convey a different meaning or message than the photograph."

Instead, the Court made clear that creating a new work that adds a "new meaning or message" does not suffice to make a use "transformative." Examining its prior decision in *Campbell v Acuff-Rose*, the court recalled:

> But new meaning or message was not sufficient. If it had been, the Court could have made quick work of the first fair use factor. Instead, meaning or message was simply relevant to whether the new use served a purpose distinct from the original, or instead superseded its objects. That was, and is, the "central" question under the first factor.[80]

The Court emphasized the significance of the commercial purpose or character of the use, recalling *Campbell v Acuff-Rose*'s sliding scale of transformativeness relative to commerciality. The Court offered a roadmap to assessing the first fair use factor:

> In sum, the first fair use factor considers whether the use of a copyrighted work has a further purpose or different character, which is a matter of degree, and the degree of difference must be balanced against the commercial nature of the use. If an original work and a secondary use share the same or highly similar purposes, and the secondary use is of a commercial nature, the first factor is likely to weigh against fair use, absent some other justification for copying.

---

[78] The Question Presented in the petition for *certiorari*, was:

    Whether a work of art is "transformative" when it conveys a different meaning or message from its source material (as this Court, the Ninth Circuit, and other courts of appeals have held), or whether a court is forbidden from considering the meaning of the accused work where it "recognizably deriv[es] from" its source material (as the Second Circuit has held).

While this presentation of the Question mischaracterized the Second Circuit's opinion, the Supreme Court appears in any event to have ignored Petitioner's framing of the issue.

[79] *Supra* note 3 at 1266 (emphasis added). See also *Andy Warhol II*, *ibid* at 1278: "the Court expresses no opinion as to the creation, display, or sale of any of the original Prince Series works."

[80] *Ibid* at 1282–1284. See also *Andy Warhol SC*, *ibid* at 1273: "Although new expression may be relevant to whether a copying use has a sufficiently distinct purpose or character, it is not, without more, dispositive of the first factor." Similarly, in confronting the tension between transformative fair use and the exclusive derivative works right, the Court stressed: "*Campbell* cannot be read to mean that §107(1) weighs in favor of any use that adds some new expression, meaning, or message." Otherwise, "transformative use" would swallow the copyright owner's exclusive right to prepare derivative works. Many derivative works, including musical arrangements, film and stage adaptions, sequels, spinoffs, and others that "recast, transfor[m] or adap[t]" the original, §101, add new expression, meaning or message, or provide new information, new aesthetics, new insights and understandings" (*Ibid* at 1282).

In shifting the focus of the first factor inquiry from the transformativeness of the defendant's *work* to the distinctness of the *use*'s purpose or character, the majority acknowledged that "The same copying may be fair when used for one purpose but not another."[81] Thus, some uses might be fair and others not, even though the uses involve the same work by the defendant. As a result, it will be important in the future to anticipate what kinds of uses for the same work will and will not be fair. For example, the decision indicates that fine art single or limited editions, as opposed to the kinds of multiples, *eg*, posters, as well as competing magazine publications, for which the plaintiff's work might also be licensed, may still be fair use. While the Second Circuit's retreat from prior Circuit caselaw suggested that court might no longer be giving high-end artists a free ride off the works of creators lower down the art world pecking order (such as photographers and graphic artists), the Supreme Court's distinctions among types of uses may still accommodate a "celebrity plagiarist privilege" for copies displayed and sold in exclusive galleries, at least where such exalted venues do not compete with the plaintiff artist's markets for selling original copies of her works.[82] Perhaps more broadly, the use-based analysis may insulate fine artists' (celebrity or otherwise) primary markets for exploitations of limited numbers of physical originals if the plaintiff's primary and secondary markets all involve licensing of mass market multiples.

One should, however, avoid concluding that the mere absence of primary market substitution will result in a defendant-favorable factor one finding; the *AWF* court also emphasized the need to justify the copying: "[a]n independent justification like this [criticism or commentary] is particularly relevant to assessing fair use where an original work and copying use share the same or highly similar purposes, *or where wide dissemination of a secondary work would otherwise run the risk of substitution for the original or licensed derivatives of it*."[83] In other words, even if the parties' exploitations differ, widespread dissemination of the defendant's work, particularly mass market exploitation, poses a risk of usurping the plaintiffs' potential derivative works markets, and therefore requires justification. Thus, because, under the court's use-focused analysis, the artist-defendant's *work* is not fair use, then even if limited gallery sales might proceed free of the underlying artist's copyright, the

---

81 *Ibid* at 1277. See also *Andy Warhol SC*, *ibid* at 1266, 1273: "the *specific use* of Goldsmith's photograph alleged to infringe her copyright is *AWF's licensing of Orange Prince to Condé Nast"* (emphasis added); *ibid* at 1278: "*In that context* [magazine publication], the purpose of the image is substantially the same as that of Goldsmith's photograph" (emphasis added); *ibid* at 128: "[The] difference [in the works] must be evaluated in the context of the specific use at issue. The use is AWF's commercial licensing of Orange Prince to appear on the cover of Condé Nast's special commemorative edition. The purpose of that use is, still, to illustrate a magazine about Prince with a portrait of Prince."). See also *Andy Warhol SC*, *ibid* at 129 (Gorsuch J, concurring): "Last but hardly least, while our interpretation of the first fair-use factor does not favor the Foundation in this case, it may in others. If, for example, the Foundation had sought to display Mr. Warhol's image of Prince in a nonprofit museum or a for-profit book commenting on 20th-century art, the purpose and character of that use might well point to fair use. But those cases are not this case."

82 See *eg*, *Graham v Prince* No 15-CV-10160 (SHS), 2023 US Dist. LEXIS 83267 (S.D.N.Y. Dist Ct, 2023) at *53 (discussing fourth factor): "[Richard] Prince has demonstrated that his 'work appeals to an entirely different sort of collector,' which lends support to the conclusion that Prince has not usurped the primary market for the original photographs."

83 *Supra* note 3 at 1263 (emphasis added in italics).

appropriating artist may not, without additional justification, necessarily exploit his work in other markets independently of the underlying artist's rights.

By contrast, if the *work* were fair use, it "[would] not [be] an infringement of copyright,"[84] and, some have argued, would enjoy its own copyright, which its author could autonomously exploit.[85] That argument, however, is problematic because it untethers the work from the reasons why it was held to be a fair use. For example, if a derivative work were ruled fair because it was created for non-commercial educational purposes, it would be perverse to contend that, once a court rules those uses of the work non infringing, its creator may now turn around and commercialize the work to the general public. Use-based analysis keeps the second work in its fair use context. That said, use-based analysis may complicate copyright claims by the second author. Suppose the Andy Warhol Foundation seeks a copyright registration for a work in the "Prince series." Under section 103(a) of the US Copyright Act:

> The subject matter of copyright as specified by section 102 includes … derivative works, but protection for a work employing preexisting material in which copyright subsists *does not extend to any part of the work in which such material has been used unlawfully*.

(emphasis added)

Whether Lynn Goldsmith's photograph was "used unlawfully" would depend on how the Foundation used her work. Incorporation of her work is not necessarily unlawful in itself, at least if confined to the display and sale of a limited number of physical originals. As to those works, it appears that the Foundation could have an infringement claim if a third party commercialized copies of the silkscreen. But the Foundation would not be entitled to license reproductions of the silkscreen in markets that compete with Goldsmith's licensing. It is not clear, under these circumstances, what disclosures the Foundation would need to make in an application for copyright registration (which is a prerequisite to bringing an action for infringement of a US work[86]). Compendium of U.S. Copyright Office Practices (2021) s 618.5 details the information a derivative works author must provide to distinguish claimed "new copyrightable material" from pre-existing content that the applicant has transformed.[87] But these specifications go to establishing authorship,

---

[84]   *Supra* note 12 at §107.

[85]   See *Keeling v Hars* 809 F 3d 43 (2nd Cir, 2015) at 49: if "a work employs preexisting copyrighted material lawfully—as in the case of a 'fair use'—nothing in the statute prohibits the extension of the "independent" copyright protection promised by 17 USC. § 103. A close reading of the statute therefore makes plain that an unauthorized but lawful fair use employing preexisting copyrighted material may itself merit copyright protection." *Cf* Abraham Bell & Gideon Parchomovsky, "Propertizing Fair Use" (2021) 107 Va L Rev 1255 at 1255: "[W]e call for the introduction of a new in rem conception of fair use, under which a fair use ruling would serve as a property remedy that shelters all subsequent users of works who fairly incorporate preexisting materials. Under this new conception, a finding of fair use would run with that new work, like an easement to all other distributors, broadcasters, publishers, performers and others who use it." These authors, however, argue that "it is not necessary for all subsequent uses to be sheltered by the initial fair use finding." *Ibid* at 1284.

[86]   *Supra* note 12, § 411(a).

[87]   See US Copyright Office, Compendium of U.S. Copyright Office Practices: Chapter 600, online: US Copyright Office <https://copyright.gov/comp3/chap600/ch600-examination-practices.pdf>.

not to limiting the uses the applicant may make of her work. Similarly, the forms for registering a derivative work call for the applicant to identify the underlying work to be excluded and to specify the new material to be included.[88] But the forms do not provide for limitations on the scope of exploitation of the copyright in the new material. If some derivative works are fair for some purposes, but not for others, it does not appear that the applicant will be able to describe the consequent scope of copyright in those works. Indeed, supposing the form allowed for it, would it even be possible to provide such a description until the copyright owner of the underlying work brings an infringement action and a court rules which, if any, uses of the derivative work are fair?[89]

Used-based analysis prompts other queries as well. What if the use adds no or little "new meaning or message," but does not compete with a market the plaintiff has exploited or seeks to exploit? Suppose the plaintiff artist does not wish to license her work for reproduction on merchandizing properties such as t-shirts and coffee mugs because she finds such uses demeaning? Does *AWF* suggest that a commercial use that the plaintiff forswears becomes "transformative" because, as to that plaintiff, it targets a non-substitutional market?[90] It may suffice to avoid that paradox by inquiring if the defendant's market is one that similarly situated creators do, or would, exploit. Lower courts have also recognized, in connection with the fourth factor, that the author need not "saturate th[]e market[]" with every conceivable derivative work in order to counter fair use claims concerning derivative works the plaintiff has not licensed.[91] Moreover, as we have seen, the defendant still must justify the use; lack of competition does not supply its own justification. Rather, courts will consider whether the defendant's work has any "critical bearing"[92] on the copied content.

The Supreme Court's de-emphasis of "new meaning or message" should not become a double-edged sword. "New meaning or message," will not carry the first fair use factor if the defendant's work lacks justification, regardless of whether it competes with a use the plaintiff has made; if the defendant's use does not compete with any of plaintiff's, it does not follow that the lack of "new meaning or message"

---

[88] See US Copyright Office, Help: Single Serial Issue, online: US Copyright Office <https://www.copyright.gov/eco/help-single-serial-issue.html#loc>.

[89] For that matter, suppose the derivative work has been produced under a license limiting the permitted exploitations, for example, the author of a novel authorizes its adaptation into a play, but does not authorize motion picture versions. If the contract transferring rights has been recorded in the Copyright Office, it might be possible to ascertain usage limitations. If the playwright creates or licenses a film version of her play, she will be infringing the copyright in the novel. In those circumstances, the film would be using the underlying novel unlawfully and any copyright in the new matter could not extend to the unlawfully used underlying content. Would the play also be using the underlying work unlawfully because the playwright's license to the filmmaker exceeded the scope of the novelist's original grant? Would the unauthorized grant of the film license therefore jeopardize the validity of the registration of the play?

[90] *Cf Twin Peaks Productions, Inc v Publications International, Ltd* 996 F 2d 1366 (2nd Cir, 1993) at 1377, (suggesting no market harm under the fourth factor when the use "filled a market niche that the [copyright owner] simply had no interest in occupying").

[91] See *eg*, *Castle Rock Entertainment, Inc v Carol Publishing Group, Inc* 150 F 3d 132 (2nd Cir, 1998) at 146: "It would ... not serve the ends of the Copyright Act—*ie*, to advance the arts—if artists were denied their monopoly over derivative versions of their creative works merely because they made the artistic decision not to saturate those markets with variations of their original."

[92] *Supra* note 3 at 1284.

becomes irrelevant to the fair use calculus, particularly if the use is commercial. Indeed, it would be perverse to contend that a commercial use that neither criticizes nor comments on the copied work – and therefore lacks justification – ought nonetheless be deemed "transformative" because it exploits a market that the plaintiff has not occupied. Such a reading would create incentives for free riders to "saturate the market" for derivative works that the first author has not or will not develop. As the district court in *Graham v Prince*[93] observed, in evaluating the fourth factor, the audiences for the works of fine artist Richard Prince and of photographer Donald Graham may not overlap, but "Prince has failed to show that other artists would not be emboldened by his success in declining to compensate plaintiffs for his non-transformative use [*ie*, a use that does not comment on or repurpose the plaintiff's expression], which negatively affects the value of original works."[94] The "exclusive Right" that the Constitution empowers Congress to "secure" to authors,[95] and the "exclusive rights" the Copyright Act confers on creators to "do or to authorize" acts within the scope of copyright (reproductions, derivative works, distributions, public performances and displays)[96] necessarily encompass the right not to do or to authorize those acts.[97] While it is true that those rights are "subject to" fair use, fair use remains an exception; an interpretation of the first factor that entitled third parties without justification, to occupy markets that the creator has not entered could effectively eviscerate the derivative works right, making fair use the rule rather than the exception.

As for *AWF*'s relevance outside the world of visual arts, *Campbell v Acuff-Rose*, on which *AWF* drew, ruled that an aesthetic transformation may not carry the first factor if it competes in a market for plaintiff's work (in that case, rap derivatives of popular songs).[98] The existence of the "arrangement" privilege in s.115[99] suggests that there are markets for all kinds of different versions of nondramatic musical works with which an unauthorized (and uncompensated) version in a different style might compete. To ascertain whether the use is a non-substitutional commentary or critical use, or a competing derivative work, the Court reaffirmed *Campbell v Acuff-Rose*'s distinction between parody and satire. Where the copied work is the object of the second work's analysis, commentary (or mockery), it is necessary to

---

[93] *Supra* note 82.

[94] *Supra* note 82 at *61.

[95] US Constitution Art I § 8 cl 8.

[96] *Supra* note 12, § 106.

[97] See *eg*, *Fox Film Co v Doyal* 286 US 123 (1932) at 127: "The owner of the copyright, if he pleases, may refrain from vending or licensing and content himself with simply exercising the right to exclude others from using his property."; *Orson, Inc v Miramax Film Corporation* 189 F 3d 377 (3rd Cir, 1999) (holding state law mandating licensing to motion picture theatres pre-empted under the Copyright Act); *Association of American Publishers, Inc v Frosh* 586 F Supp 3d 379 (D.Md Dist Ct, 2022) (holding state law mandating licensing eBooks to public libraries pre-empted under the Copyright Act).

[98] *Supra* note 3 at 1275, citing *Campbell*, *supra* note 27 at 580–583.

[99] *Supra* note 12, § 115(a)(2) provides:

   Musical arrangement. —A compulsory license includes the privilege of making a musical arrangement of the work to the extent necessary to conform it to the style or manner of interpretation of the performance involved, but the arrangement shall not change the basic melody or fundamental character of the work, and shall not be subject to protection as a derivative work under this title, except with the express consent of the copyright owner.

copy as much as needed to support the commentary. By contrast, adhering to a dis-tinction the Court of Justice of the European Union ("CJEU") has rejected,[100] the Court emphasized that "[p]arody needs to mimic an original to make its point, and so has some claim to use the creation of its victim's (or collective victims') imag-ination, whereas satire can stand on its own two feet and so requires justification for the very act of borrowing."[101] When the uses of works compete, for example, both exploit commercial markets for popular music, the defendant must provide a compelling justification for why its "copying is reasonably necessary to achieve the user's new [critical] purpose."[102] In the *AWF* case, "because AWF's commercial use of Goldsmith's photograph to illustrate a magazine about Prince is so similar to the photograph's typical use, a particularly compelling justification is needed. Yet AWF offers no independent justification, let alone a compelling one, for copying the pho-tograph, other than to convey a new meaning or message. As explained, that alone is not enough for the first factor to favor fair use."[103]

Two examples may illustrate the likely application of *AWF* to works outside the visual arts. In *Suntrust Bank v Houghton Mifflin Co*,[104] the owner of the copyright in the romance of the antebellum South, *Gone With The Wind* ("*GWTW*"), brought an infringement action against the publisher of the novel *The Wind Done Gone*, which retold portions of the prior work from the point of view of the people enslaved by the first novel's protagonists. While the plaintiff asserted the defendant's novel was an unauthorized sequel, the publisher prevailed on its fair use defense that the sec-ond work criticized and subverted the first work's white supremacist sympathies. Both works were novels, and at a superficial level, competed in the same market, but the purpose of the defendant's novel was "to rebut and destroy the perspective, judgments, and mythology of *GWTW*. [The defendant's] literary goal is to explode the romantic, idealized portrait of the antebellum South during and after the Civil War."[105] Using the characters and plot elements of *GWTW* to turn the iconic novel against itself and publicly expose its deep-seated racism provided a "compelling justification" for the copying. Indeed, the *AWF* Court referenced the "Wind Done Gone" case as an example of justified copying.[106]

By contrast, in *Dr. Seuss Enters., L.P. v ComicMix LLC*,[107] the Ninth Circuit rejected the fair use defense in a case concerning a combination of elements of two different works: the illustrated book *Oh, the Places You'll Go!* ("*Go!*"), by children's author "Dr. Seuss," and the popular 1960s television series *Star Trek*, into an illustrated book titled *Oh the Places You'll Boldly Go!* ("*Boldly*")[108] The

---

[100] *Deckmyn v Vandersteen* Case C-201/13, [2014] ECLI:EU:C:2014:2132 at ¶ 33 (holding that "the con-cept of 'parody' . . . is not subject to the condition[] that the parody . . . relate to the original work itself or mention the source of the parodied work").

[101] *Supra* note 3 at 1275, citing *Campbell*, *supra* note 27 at 580, 581.

[102] *Ibid*.

[103] *Ibid* at 1285, 1286.

[104] *SunTrust Bank v Houghton Mifflin Co* 268 F 3d 1257 (11th Cir, 2001).

[105] *Ibid* at 1270.

[106] See *Andy Warhol SC*, *supra* note 3 at 1285 n 21.

[107] *Supra* note 61.

[108] The mashup's title borrowed from the opening words of each episode "Space, the final frontier/These are the voyages of the Starship Enterprise/Its five year mission/To explore strange new worlds/To seek

court observed that the defendant's work neither critiqued nor commented on the plaintiff's, nor mocked its author's style. Rather, "ComicMix wanted *Boldly* to be a *Star Trek* primer that 'evoke[s]' rather than 'ridicule[s] *Go!*.'"[109] Moreover, *Go!* had spawned multiple licensed derivatives, and *Boldly* was created to compete in the market for derivative works based on *Go!* In that case, defendant's work directly competed in the plaintiff's derivative markets, and the defendant offered no justification other than its provision of "extensive new content."[110] The Ninth Circuit's observation that "[a]bsent new purpose or character, merely recontextualizing the original expression by 'plucking the most visually arresting excerpt[s]' of the copyrighted work is not transformative,"[111] remains equally viable after *AWF*, as does the appellate court's caution that "But the addition of new expression to an existing work is not a get-out-of-jail-free card that renders the use of the original transformative."[112] In sum, nothing in *AWF* suggests different outcomes for these cases, one involving a non-substitutional critique, the other a commentary-free mashup of work that had extensively exploited the derivative works market.

III. Fair use and AI Inputs

I turn now to a more speculative analysis, of fair use and artificial intelligence. AI systems such as ChatGPT for generating literary works and Dall-E and MidJourney for generating images, have lately attracted considerable attention, and consternation. Their outputs may convincingly resemble human-made text and images, even though the extent of human involvement in the creation of the output may be quite scant, limited to a few text prompts. While much has been written regarding the copyrightability of AI-generated (or assisted) outputs,[113] my present inquiry concerns those systems' inputs. For an AI system to generate text, images, music, or computer code, it must "learn" from vast databases of literary, artistic or musical works. If the source content appeared on platforms whose terms and conditions require the posting authors to permit off-platform uses, including reproduction, modification, and further communication of the works, then no question of infringement

---

out new life/And new civilizations/**To boldly go** where no man has gone before" (emphasis added). The producers of Star Trek were not parties to the lawsuit.

[109] *Supra* note 61 at 453.

[110] *Ibid* at 453.

[111] *Ibid* at 454.

[112] *Ibid* at 453.

[113] See *eg*, US Copyright Office, Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, online: US Copyright Office <https://www.copyright.gov/ai/ai_policy_guidance.pdf>; Vicenç Feliú, "Our Brains Beguil'd: Copyright Protection for Ai Created Works" (2021) 25(2) Intell Prop & Tech L J 105; Ginsburg & Budiardjo, *supra* note 4; Nina I. Brown, "Artificial Authors: A Case for Copyright in Computer-Generated Works" (2018) 20(1) Colum Sci & Tech L. Rev 1; Kalin Hristov, "Artificial Intelligence and the Copyright Dilemma" (2017) 57(3) IDEA 431 at 441; James Grimmelmann, "There's No Such Thing as a Computer-Authored Work—And It's a Good Thing, Too" (2016) 39(3) Colum J L & Arts 403; Annemarie Bridy, "Coding Creativity: Copyright and the Artificially Intelligent Author" [2012] 2012 Stan Tech L Rev 5; Pamela Samuelson, "Allocating Ownership Rights in Computer-Generated Works" (1986) 47(4) U Pitt L Rev 1185.

arises because the subsequent uses will have been authorized.[114] In the absence of authorization, if the works constituting the system's "training data" are under copyright, does their incorporation into computer memory infringe the copyrights in the uploaded works?

Copying entire works and permanently or temporarily storing them in computer memory would constitute prima facie infringement. But an applicable exception may excuse what would otherwise be infringing conduct. The US Copyright Act[115] includes a variety of specific exemptions or limitations, but none appear to correspond to massive wholesale copying for the purpose of creating training data.[116] That leaves fair use: do that defense's contours accommodate this use? Fair use issues may arise both at the input (or ingestion) stage, as well as with respect to the outputs.[117] While I will focus primarily on the input stage, potentially pertinent caselaw couples

---

[114] Where the platform imposes terms and conditions regarding its use, such as attribution or share-alike, then failure to abide by those conditions could result in a breach of a material term of the license, leading to a finding of infringement, see *eg*, *Software Freedom Conservancy, Inc v Vizio, Inc*, No 8:21-cv-01943-JLS-KES, 2022 US Dist. LEXIS 87115 (C.D. Cal. Dist Ct, 2022) (breach of material term of General Public License held to infringe freely available software); *Artifex Software v Hancom, Inc*, No 16-cv-06982-JSC, 2017 US Dist. LEXIS 62815 (N.D. Cal. Dist Ct, 2017); *Jacobsen v Katzer* 609 F Supp 2d 925 (N.D. Cal. Dist Ct, 2009), dismissed in *Jacobsen v Katzer* 449 F App'x 8 (Fed Cit, 2010). A lawsuit filed against AI producer Github for failing to abide by the attribution condition term of the GPL initially survived a motion to dismiss, see *Doe 1 v Github*, No 22-CV-06823-JST, 2023 US Dist. LEXIS 86983, (N.D. Cal. Dist Ct, 2023); those claims were subsequently dismissed, see *DOE 1 v GitHub*, No 4:22-CV-06823, (N.D. Cal. Dist Ct, 2024) ECF No. 189; that dismissal is now under reconsideration. Many UGC platforms state grants of license from users in terms sufficiently broad to cover AI data training by the platform or its licensees. See *eg*, X (formerly Twitter): "use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such [user] Content in any and all media or distribution methods now known or later developed": Your Rights and Grant of Rights in the Content, X: Terms of Service <https://twitter.com/en/tos>. Some licenses specifically authorize AI uses. See *eg*, Google: general terms of use of large platforms such as Google: "license […] includes using automated systems and algorithms to analyze your content […] to recognize patterns in data." Your relationship with Google, online: Google: Privacy & Terms <https://policies.google.com/terms?hl=en-US#toc-relationship>. Image generation sites may require users to allow the incorporation of user-requested outputs into the training data. See *eg*, Midjourney: "We may use Your information for other purposes, such as data analysis, identifying usage trends, … and to evaluate and improve our Service…" *Privacy Policy*, online: Midjourney: Documentation <https://docs.midjourney.com/docs/privacy-policy>; ChatGPT "improves by further training on the conversations people have with it, unless you choose to disable training" *How your data is used to improve model performance*, online: OpenAI: Help <https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance>.

[115] *Supra* note 12.

[116] Compare Directive 2019/790, of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market, Arts 3 and 4, establishing exceptions for text and data mining.

[117] Because fair use is an affirmative defense to infringement, predicate questions inquire whether at the input stage the system has made a substantial, non-transient copy of a copyrighted work; and whether the outputs reproduce recognizable portions of the copied works. For present purposes, I will assume that the training data incorporates reproductions of multiple works of authorship. See Niv Haim *et al*, Reconstructing Training Data from Trained Neural Networks (2022), https://arxivorg/abs/2206.07758 [Haim]; Nicholas Carlini *et al*, "Extracting Training Data from Large Language Models" presented at USENIX Security Symposium (11-13 August 2021) <https://arxivorg/abs/2012.07805>. *Cf* Rosana Ducato *et al*, "Ensuring Text and Data Mining: Remaining Issues with the EU Copyright Exceptions and Possible Ways Out" (2021) Eur IP Rev 322 at 334: "The expressive features of the work are not used, and there is no public to enjoy the work, as the work is only an input in a process for searching a corpus and identifying occurrences and possible trends or patterns"; Matthew Sag, "The New Legal Landscape for Text Mining and Machine Learning" (2019) 66(2) J. Copyright Soc'y U.S.A. at 300.

the inputs to the outputs, excusing the former if, among other considerations, they are necessary to the production of non-infringing outputs. The leading example of these cases, *Authors Guild v Google, Inc*[118] concerned the mass digitization of millions of in-copyright books in the University of Michigan library in order to constitute a database that could be searched for information about, and limited excerpts ("snippets") from, the copied books. Arguably, copying works to enable AI systems to "learn" how to produce independent outputs consisting of literary, artistic, musical, audio-visual works or software, sufficiently repurposes the copying to count as "transformative"[119] – at least if the outputs enabled by the inputs do not themselves infringe the source content[120] (a point of considerable contention[121]).

---

[118] *Authors Guild v Google, Inc* 804 F 3d 202 (2nd Cir, 2015) [*Authors Guild v Google*].

[119] *Cf supra* note 3 at 1277 n 8, referencing *supra* note 2: "Google put Sun's code to use in the 'distinct and different computing environment' of its own Android platform, a new system created for new products." The highly functional nature of the plaintiff's work in *Google v Oracle* nonetheless calls for caution in citing its statements out of their particular context."

[120] See *Authors Guild v Google*, *supra* note 118; *supra* note 108 (copying and retention of millions of copyrighted books into search engine database held fair use when the outputs of searches were limited to uncopyrightable bibliographic information or nonsubstitutional "snippets").

[121] Several recently filed lawsuits contend that the outputs incorporate the expression of the copied works, see eg, *Nazemian v NVIDIA Corp, No 3:2024cv01454* (filed 8 March 2024) (N.D. Cal.); *Basbanes v Microsoft Corporation* No 1:24-cv-00084 (filed January 5, 2024) (S.D.N.Y.); *The New York Times Company v Microsoft Corporation* No 1:23-cv-11195 (filed 27 December 2023) (S.D.N.Y.); *Concord Music Group, Inc v Anthropic PBC* No 3:23-cv-01092 (filed 18 October 2023) (M.D. Tenn.); *J.L. v Alphabet Inc* 3:23-cv-03440 (filed 11 July 2023) (N.D. Cal.); *Kadrey v Meta Platforms, Inc* No 3:23-cv-03417, US Dist. LEXIS 207683 (Dist Ct.) (filed 12 February 2023) (filed 12 February 2023) motion to dismiss granted; *Kadrey v Meta Platforms, Inc* No 3:23-cv-0341 (N.D. Cal. Dist Ct, 2023) ECF No 56, amended complaint filed, alleging direct copyright infringement through unauthorized copying of plaintiffs' books for purposes of training LLaMA models; *Getty Images v Stability AI, Inc* No 3:23-cv-00201 (filed 3 February 2023) (D. Del.); *Getty Images (US), Inc v Stability AI, Inc* No 1:23-cv-00135-UNA (filed 3 February 2023) (D. Del.); *Andersen v Stability AI Ltd* No 3:23-cv-00201 (filed 13 January 2023) (N.D. Cal.) [*Anderson v Stability AI Ltd*] (Dismissed in full, except for the copyright infringement claim against Stability AI, with leave to amend by November 29, 2023 (ECF No 117 at 28.4, 28.5). Note that the claim against Stability AI was based on its unauthorized use of copyrighted images to create Stable Diffusion, as well as the storage of those images on the software—not on the infringing status of the program's outputs (*Ibid* at 7.8–7.12). In anticipation of the amended complaint, the court appraised that defendants made a "strong case" to dismiss the argument that output images are derivative works "because plaintiffs cannot plausibly allege the Output Images are substantially similar or re-present protected aspects of copyrighted Training Images, especially in light of plaintiffs' admission that Output Images are unlikely to look like the Training Images" (*Ibid* at 13.1–13.4). However, the court is still considering the novel proposition that "Output Images can be so similar to plaintiff's styles or artistic identities to be misconstrued as "fakes"" (*Ibid* at 13.5–13.6). The amended complaint reasserts a copyright infringement claim based on the substantial similarity of out¬puts to training data (ECF No 129). Moreover. studies concerning AI copying of news publisher content, undertaken subsequently to the pleadings in *Andersen v Stability AI Ltd*, indicate a high degree of trace¬ability of publisher content in the outputs of programs such as Chat GPT. See *eg*, Nick Diakopoulus, "Finding Evidence of Memorized News Content in GPT Models", *Medium* (5 September 2023); US Copyright Office Inquiry Artificial Intelligence and Copyright, Comments of News Media Alliance, Appendix A White Paper: How the pervasive copying of expressive works to train and fuel generative artificial intelligence systems is copyright infringement and not a fair use, Technical Appendix, part 3. Similarities Between Publisher Content and Long-Form LLM Outputs, online: regulations.gov <https:// www.regulations.gov/comment/COLC-2023-0006-8956>.

At the hearing for the Defendants' Motion to Dismiss in *Andersen v Stability AI Ltd*, the Court stated it was "inclined to dismiss almost everything [in the Complaint] with leave to amend," Lovejoy Decl.Ex. A (Hearing Tr.). 4:10–6:6, Aug. 2, 2023. ECF No 92. In the meantime, the Court has scheduled a Case Management Conference for Oct. 10, 2023. Clerk's Notice Resetting Zoom Hearing, Sept. 13, 2023. ECF No 111).

1. *Fair use analysis in light of an emerging market for licensing training data*

But one should perhaps decouple the inputs from the outputs. Looking only at whether the copying of works into training data is a "transformative" fair use, *AWF* suggests that analysis may depend on whether there is a market for licensing content for training data. Such markets do exist, notably in news media, for high quality, reliable data,[122] and other authors and copyright owners are endeavoring to develop those markets as well.[123] In that event, even if the outputs might not infringe par-

---

[122] See *eg*, Gerrit de Wynck, "OpenAI strikes deal with AP to pay for using its news in training AI", *Washington Post* (13 July 2023). The Internet forums Reddit and Stack Overflow plan to begin charging AI developers to use text from their sites as training data for algorithms, see *eg*, Paresh Dave, "Stack Overflow Will Charge AI Giants for Training Data", *Wired* (20 April 2023); Mike Isaac, "Reddit Wants to Get Paid for Helping to Teach Big A.I. Systems", *New York Times* (18 April 2023). Shutterstock, which licensed stock image datasets to train OpenAI's text-to-image generator DALL-E, continues to offer tagged dataset packages through Amazon Web Services' Data Exchange platform. Shutterstock Seller Profile, online: AWS Data Exchange <https://aws.amazon.com/marketplace/seller-profile?id=f-b34254c-c7cf-47b8-806c-24045a0a2807>. See also Comments submitted before the U.S. Copyright Office's Notice of Inquiry on Artificial Intelligence and Copyright in U.S. Copyright Office, Artificial Intelligence and Copyright, online: regulations.gov <https://www.regulations.gov/docket/COLC-2023-0006/comments> [Copyright Office NOI]. "Only a select group of publishers, primarily the largest brands with premium content, established audiences, and robust infrastructure, have the privilege to capitalize on their works through paywalls and licensing. Those outside this echelon will curtail their production of news, and many will simply go out of business" (News Corporation); "Companies' infringing uses of our content… reduce the value of, and harm, our existing licensing business, which grants licenses for a wide variety of uses, including data mining and media monitoring" (The New York Times)). Unmediated "scraping" of the internet in order to generate training data can produce unreliable data models, see *eg*, Robert McMillan, "AI Junk Is Starting to Pollute the Internet", *Wall Street Journal* (12 July 2023). For an introduction to data marketplaces and the obstacles to establishing an enduring exchange platform, see Markus Spiekermann, "Data Marketplaces: Trends and Monetisation of Data Goods" (2019) 54 Intereconomics 208.

[123] See *eg*, *Authors Guild v OpenAI Inc*, Case 1:23-cv-08292 (filed 19 September 2023) (D. NY) [*Authors Guild v OpenAI*] (Class of professional fiction writers seek, among other remedies, damages for "the lost opportunity to license their works", which were used as training data without their permission); Brian Fung, "Thousands of authors demand payment from AI companies for use of copyrighted works", *CNN* (20 July 2023); "Survey Reveals 90 Percent of Writers Believe Authors Should Be Compensated for the Use of Their Books in Training Generative AI", *The Authors Guild* (15 May 2023). Cecily Mauran, "Google and Universal Music might license artists' voices for AI-generated music", *Mashable* (10 August 2023). "A New A.I. Image Generator Is Promising to Pay Royalties to Artists Who Submit Work to Train Its Model", *Artnet News* (19 May 2023); Emilia David, "Getty made an AI generator that only trained on its licensed images", *The Verge* (25 September 2023). See also Copyright Office NOI, *ibid*. ("Most copyright owners have recognized the value of generative AI licensing and have developed, or are in the process of developing, licensing models" (Copyright Alliance); "UMG licenses its content for various technological and AI-powered purposes, including for example, for the development of fingerprint systems and music recommendation systems. While the existing market is vast, the potential market is unlimited, and AI training is the latest iteration of an emerging but very real market. Allowing training without license essentially destroys that market before it can mature" (Universal Music Group); "At this time, there are no known licensing deals between book publishers and AI developers, but we know that some companies have started to reach out to trade (commercial) publishers to request licenses and some trade publishers are considering them" (Authors Guild); "In the past 12 months we have heard more about AI companies licensing images for use in future training sets, so we asked the members if they had entered into such an agreement. Exactly four of the 156 respondents (2.5%) had entered into this type of agreement" (American Society of Media Photographers and North American Nature Photographers Association)).

ticular inputs, commercial copying (at least) to create training data would be for the same purpose, and might therefore fail a first factor fair use inquiry after *AWF*, without a "compelling justification" for supplanting authors' markets. Moreover, unlike the source copies for Google's book-scanning program, which were lawfully acquired library books, the source copies "scraped" from the Web, or from a corpus of books or other content, may be infringing.[124] The compiler of the training data's knowledge of the unlawful provenance of the source copies might well taint the "character" of the defendant's use.[125]

Assuming that taint does not disqualify the fair use claim, or that the source copies were lawfully made, the justification in this instance might emphasize a different kind of purpose of the use; rather than considering whether the use competes with the licensing of training data (which, we are positing, it does), the fair use proponent would urge that the purpose of the use is not to exploit the copied works' expressive qualities, but instead merely to dismember the copied content into indicia of how works are constructed.[126] In other words, the use does not target the copied content as a work, but rather as data evidencing language use, image composition, musical note sequencing, etc. Fair use may not absolve copying for the same expressive purpose; it arguably will shield copying that merely provides information about the copied works,[127] whether the recipient of that information is a human or a machine.

This argument, however, depends on the equation of humans and machines, which may be a false equivalence. The "information about the works" that systems

---

[124] See *Authors Guild v Open AI*, *supra* note 123 (contending that OpenAI trains on corpora of pirated books).

[125] See *Harper & Row, supra* note 12 at 563 (underscoring the defendant's preparation of its work from a "purloined manuscript"); Lloyd Weinreb, "Fair's Fair" (1990) 103 Harv L Rev 1137 at 1152: "[I] t makes a difference whether a user obtained his copy of the original work lawfully or by theft, and if lawfully, by a means that is entirely proper or in some manner underhanded. It makes a difference whether a copyright owner's reason for refusing to give a license for the use is one that the community generally approves, copyright issues aside, or is one that it allows but disapproves."). For case law, see *eg*, *F. Marc Schaffel Productions, LLC v TMZ Productions* No CV 10-01306 GHK (SSx), 2010 US Dist. LEXIS 151990 (C.D. Cal. Dist Ct, 2010) at *9; *TELUS Corp v Watson* No C 07-3434 VRW 2009 US Dist. LEXIS 152131, at *7 (N.D. Cal. Dist Ct, 2009); *Los Angeles News Service v KCAL-TV Channel 9* 108 F 3d 1119 (9th Cir, 1997) at 1122; *DSC Communication Corp v DGI Technologies Inc* 898 F Supp 1183 (Dist Ct, 1995) at 1194, affirmed in *DSC Communication Corp v DGI Technologies, Inc* 81 F 3d 597 (5th Cir, 1996); *Atari Games Corp v Nintendo of America, Inc* 975 F 2d 832 (Fed Cit, 1992) at 843; *Radji v Khakbaz* 607 F Supp 1296 (D.D.C. Dist Ct, 1985) at 1300, 1301, amended, No CIVA. 84-0641, 1987 WL 11415 (D.D.C. May 15, 1987); *Marcus v Shirley Rowley & San Diego Unified School. District* 695 F 2d 1171 (9th Cir, 1983) at 1175, 1176.

[126] See *eg*, Ducato, *supra* note 117 at 334: "The expressive features of the work are not used, and there is no public to enjoy the work, as the work is only an input in a process for searching a corpus and identifying occurrences and possible trends or patterns"; Matthew Sag, *supra* note 117. The proposition that AI programs do not copy works for their expressive content, however, is highly contested. Benjamin L. W. Sobel, "Artificial Intelligence's Fair Use Crisis" (2017) 41(1) Colum. JL & Arts 45 at 57: "Instead of merely deriving facts about a work, [computers] may be able to glean value from a work's expressive aspects; as a result, these uses of machine learning may no longer qualify as non-expressive in character". Even scholars arguing that most machine learning systems do not copy works for their expressive elements acknowledge that some do. Mark A. Lemley and Brian Casey, "Fair Learning" (2021) 99(4) Tex L. Rev 743 at 766, 767.

[127] See *Authors Guild v Google*, *supra* note 118.

like Google Books deliver to human beings focuses on the copied book as a whole, enabling the user to make judgments about the relevance of the referenced book. By contrast, the "information" that the AI system receives is not "about" the book as a work of authorship (indeed, that is the crux of the "work as data" contention[128]), and, as we shall see, what the AI system delivers to the end-user is not necessarily "about" the copied works either.

The argument also depends on the existence of a clear line between "information about" and expression. A complete disaggregation of a work's expression – for example, into the arrangement of letters, their ordering, the rules of grammar, etc. – provides information derived from the work, but, appropriately prompted, that information permits easy reconstruction (or "regurgitation") of the work as a whole or of substantial portions of it.[129] Unless, as in *Google Books*, the program's design prevents the user from reassembling substantial portions of the work's expression,[130] and disables the program from reconstructing the expression *sua sponte*, the information/expression distinction may be untenable.

Consider the following exercise using the DALL-E image generation program. The exercise assumes that DALL-E is not storing complete source images in its training data, but has in fact disaggregated the images into informational components. When prompted "Hello, Kitty," the program delivered four images of kittens:

But when prompted "Hello Kitty at Tokyo Tower," the program returned four images nearly identical to the visual character, demonstrating that the program is capable of reassembly of the information into copies or close imitations of the source works.

---

128 See *eg*, Lemley & Casey, *supra* note 126. "Similarly, a natural language generation system wants to see what you wrote to learn how words are sequenced in ordinary conversation, not because it finds your prose particularly expressive or because it wants to use your turn of phrase."

129 See Peter Henderson *et al*, "Foundation models and fair use", online: ᴀʀXɪᴠ <https://arxivorg/abs/2303.15715>; Niv Haim *et al*, "Reconstructing Training Data from Trained Neural Networks", online: ᴀʀXɪᴠ <https://arxivorg/abs/2206.07758>; Carlini *et al*, *supra* note 117. But see *OpenAI and journalism* (Jan. 8, 2024), online: OpenAI <https://openai.com/blog/openai-and-journalism> (OpenAI acknowledges instances of "regurgitation" and claims "The New York Times is not telling the full story").

130 *Authors Guild v Google*, *supra* note 118 at 223: "As snippet view never reveals more than one snippet per page in response to repeated searches for the same term, it is at least difficult, and often impossible, for a searcher to gain access to more than a single snippet's worth of an extended, continuous discussion of the term."

Similarly, New York Times journalists obtained an image closely resembling Super Mario when prompting AI to "create a videogame plumber".[131]

AI companies claim that they have set up guardrails to prevent reproductions of copyrighted content. To the extent copyrighted material slips through, this is a "bug…they are trying to fix".[132] Professor and AI expert Gary Marcus has teamed up with artist Reid Southern and Professor Katie Conrad to test MidJourney and DALL-E's guardrails. They found that prompts to MidJourney often generate near-identical images to copyrighted movie stills and video game characters.[133] DALL-E created artwork closely resembling Super Mario when prompted to create "an Italian video game character".[134] Similarly, New York Times journalists found that ChatGPT created an image strongly resembling SpongeBob SquarePants when prompted to do so. ChatGPT generated an image even closer to SpongeBob when prompted to create "an animated sponge wearing pants".[135]

Moreover, competition with a market for training data remains, even if the compiler of the dataset copies for a merely informative purpose, and disables outputs that convey the source works' expression. Indeed, the point is to provide data for decomposition into training sets. To break out of the circle of market competition, one might contend, as a normative matter, that there is no market for uses

---

[131] Stuart Thompson, "We Asked A.I. to Create the Joker. It Generated a Copyrighted Image", *The New York Times* (25 January 2024).

[132] *Ibid*.

[133] Gary Marcus & Reid Southern, "Generative AI Has a Visual Plagiarism Problem" (6 January 2024), online: The Institute of Electrical and Electronics Engineers <https://spectrum.ieee.org/midjourney-copyright>.

[134] Gary Marcus & Katie Conrad, "DALL-E's New Guardrails: Fast, Furious, and Far from Airtight" (10 January 2024), posted in *Marcus on AI*, online: Substack <https://garymarcus.substack.com/p/dall-es-new-guardrails-fast-furious.>

[135] Thompson, *supra* note 131.

that treat works as collections of information-yielding data.[136] The Supreme Court in *Campbell v Acuff-Rose* adopted a similar approach regarding parodies when it declared "there is no protectible derivative market for criticism."[137] Thus, even if some parodists obtained licenses,[138] that practice should not count to evidence a market,[139] lest copyright owners invoke that market to silence criticism.

Transposing the normative approach to the issue of copying to assemble training data, the question becomes whether courts should not take into account the emergence of a market for licensing training data because giving copyright owners control over that market will suppress the development of AI and its attendant social benefits. In *Campbell v Acuff-Rose*, the Supreme Court posited that copyright owners would not want to license parodies: "the unlikelihood that creators of imaginative works will license critical reviews or lampoons of their own productions removes such uses from the very notion of a potential licensing market."[140] In the case of training data, by contrast, we do not confront a prospective refusal to license; unlike criticism-wary creators, copyright owners by and large are not seeking to prevent the development of AI;[141] they want to be paid for the use of their content.[142] It is not apparent, therefore, why a market for training data should not, as a normative matter, exist. Nor is it clear that requiring payment for training data will in fact suppress or retard advances in AI technology.[143] If – and it may

---

[136] Compare *Authors Guild v Google*, *supra* note 118 at 207: "an author's derivative rights do not include an exclusive right to supply information (of the sort provided by Google) about her works." The use deemed fair in Google Books digitized books in order to provide limited amounts of expression to inform users *about the copied books*, not for purposes unrelated to the contents of the copied books.

[137] *Campbell*, *supra* note 27 at 592.

[138] See *eg*, FAQ, online: "Weird Al" Yankovic: The Official Website <http://www.weirdal.com/faq.htm> : "While the law supports his ability to parody without permission, [Weird Al] feels it's important to maintain the relationships that he's built with artists and writers over the years. Plus, Al wants to make sure that he gets his songwriter credit (as writer of new lyrics) as well as his rightful share of the royalties."

[139] *Campbell*, *supra* note 27 at 592: "to the extent that the opinion below may be read to have considered harm to the market for parodies of "Oh, Pretty Woman," . . . the court erred."

[140] *Ibid*. The Court might well also have observed that "licensed parody" is an oxymoron.

[141] For example, artists Greg Rutkowski, who is known for his distinctive fantasy landscapes, and whose name has been used to generate over 90 thousand images on Stable Diffusion, commented that: "I'm not against the AI overall, I think it's a good technology. But I think they should have excluded [living] artists' names from the program," he said. Beatrice Nolan, "Artists say AI image generators are copying their style to make thousands of new images — and it's completely out of their control", *Business Insider* (17 October 2022).

[142] See *eg*, *Authors Guild v OpenAI*, *supra* note 123; Fung, *supra* note 115; The Authors Guild, *supra* note 123. Shutterstock, which has integrated OpenAI's DALL-E text-to-image generator to supplement its repository of stock imagery, launched the Contributor Fund to reimburse artists whose works are being used in the process. James Vincent, "Shutterstock will start selling AI-generated stock imagery with help from OpenAI, *The Verge* (25 October 2022). *Cf Stewart v Abend* 495 US 207 (1990) at 228 (regarding the scope of a renewal rights reversion): "Presumably, respondent is asking for a share in the proceeds because he wants to profit from the distribution of the work, not because he seeks suppression of it."

[143] For the claim that the transaction costs of paying authors will compromise technological development without actually helping authors, see *eg*, Lemley & Casey, *supra* note 126 at 748: "Because training sets are likely to contain millions of different works with thousands of different owners, there is no plausible option simply to license all of the underlying photographs, videos, audio files, or texts for the new use. So allowing a copyright claim is tantamount to saying, not that copyright owners will get paid, but that

be a big "if" – administratively reasonable means of licensing critical masses of content exist, then the transaction costs that might otherwise hobble development of beneficial new forms of exploiting works of authorship no longer appear insuperable, and there may be no other reason, practical or normative, for amputating this emerging market from copyright owners' control. As the Second Circuit recognized in *American Geophysical Union v Texaco*,[144] concerning fair use, photocopying and market substitution under the fourth factor:

> the fourth factor will favor the secondary user when the only possible adverse effect occasioned by the secondary use would be to a potential market or value that the copyright holder has not … reasonably been able to[] obtain or capture. *Pacific and Southern Co. v Duncan* 744 F.2d 1490, 1496 (11th Cir, 1984) (noting that the fourth factor may not favor copyright owner when the secondary user "profits from an activity that the owner could not possibly take advantage of").

> …

> [The publishers] have created, primarily through the [collective licensing mechanism of the] C[opyright] C[learance] C[enter], a workable market for institutional users to obtain licenses for the right to produce their own copies of individual articles via photocopying. The District Court [observed that "in this manner, private cooperative ingenuity has found practical solutions to what had seemed unsurmountable problems." 802 F.Supp. 1, 25 (S.D.N.Y. 1992)][145]

AWF makes licensing a matter of both first and fourth factor analysis, dividing the inquiry into the existence of a licensing market in the first place, and the effect of the unlicensed use on the market(s) for the work in the fourth. The viability of the licensing market would seem now to be a first factor question; *AWF* looked to whether the defendant's use substituted for the plaintiff's exploitation. If the plaintiff (or a similarly situated plaintiff) is not "reasonably able to capture" that market, then, perhaps, there is currently no market for which the defendant's use substitutes, although the present development of licensing markets in at least some sectors[146] indicates otherwise.[147] If a normatively-based objection to paying for training data

---

the use won't be permitted at all"; Dave, *supra* note 114 : "Having to pay for data they once grabbed for free could extend the already unclear timelines to turning a profit on their emerging technologies."

[144] *American Geophysical Union v Texaco* 60 F 3d 913 (2nd Cir, 1994) [*American Geophysical Union*].

[145] *Ibid* at 930. See also *American Geophysical Union*, *ibid* at 930, 931: "Despite Texaco's claims to the contrary, it is not unsound to conclude that the right to seek payment for a particular use tends to become legally cognizable under the fourth fair use factor when the means for paying for such a use is made easier. This notion is not inherently troubling: it is sensible that a particular unauthorized use should be considered "more fair" when there is no ready market or means to pay for the use, while such an unauthorized use should be considered "less fair" when there is a ready market or means to pay for the use. The vice of circular reasoning arises only if the availability of payment is conclusive against fair use."

[146] See *eg* cites, *supra*, 122, 123.

[147] One should take care, lest this analysis prejudice authors whose entrepreneurial efforts fall short, where others, more adept at marketing, might succeed.

fails to persuade, then *AWF* suggests that much may turn on the practical ability to develop an efficient licensing market for training data.

2. *Fair use analysis in the absence of a market for licensing training data*

In the absence of a recognized market for training data, the fair use status of the inputs may well depend on whether the outputs infringe. The Second Circuit's 2015 decision in *Authors Guild v Google*[148] underlies arguments that a capacious concept of transformativeness might shelter the creation of training data. We first will examine and apply that decision, and then will consider some distinguishing factors that might prompt a different conclusion.

In a decision that in many ways presaged *Authors Guild v Google*, the Second Circuit in *Authors Guild v HathiTrust* ("*HathiTrust*"),[149] concerning non-profit university library uses of their holdings as digitized by Google, found the inputting of full copies of in-copyrighted books by scanning and permanently storing them in a database to further the "transformative use" of allowing "data mining" of the contents of the books. Such uses produce no new expression by the copying and storage entities, and the outputs enabled by the "mining" of the scanned book seek not to expose its expression, but rather to extract information, for example concerning frequency of word use, or a date range when, in a corpus of millions of books, particular words or phrases first appear.

In *Authors Guild v Google*, a for-profit corporation scanned millions of in-copyright books and permanently stored their full contents in its database, thus enabling user datamining of the scanned corpus. Unlike Hathi Trust, however, Google also communicated "snippets" of the stored text in response to search queries in order to enable the user to determine the relevance of the book to her research. The Second Circuit accorded scant weight to the commercial nature of Google's enterprise, stressing that it has "repeatedly rejected the contention that commercial motivation should outweigh a convincing transformative purpose and absence of significant substitutive competition with the original."[150]

Distinguishing between outputs that convey information *about* the scanned book from outputs that convey its *expression*, the *Authors Guild v Google* court ruled that neither the datamining uses nor the snippet views exploited the copied works for their expressive value. Hence "the [inputting] of complete digital copies of copyrighted works [results in] transformative fair uses when the copies 'served a different function from the original.'"[151] While the output of snippet views did convey limited amounts of expression, the search program did not permit assembly of substantial amounts of text through reiterated sequential snippet requests. The court repeatedly emphasized the very constrained and controlled, "fragmentary and scattered," "cumbersome, disjointed, and incomplete nature of the aggregation of

---

[148] See *Authors Guild v Google* (*per* Leval J), *supra* note 118.

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

[150] *Ibid* at 219. This statement remains valid after, *AWF*: the Supreme Court stated that the commercial character of the use weighs more heavily the less transformative the purpose. See *supra* note 2 at 1263.

[151] *Authors Guild v Google*, *supra* note 118 at 217.

snippets made available through snippet view."[152] As a result, and endeavoring to avoid slippery-slope expansion of the content or presentation of fair use-permissible snippets, the Second Circuit emphasized that "at least as presently structured by Google, the snippet view does not reveal matter that offers the marketplace a significantly competing substitute for the copyrighted work."[153]

Whether the expression that appears in AI outputs amounts to a "significantly competing substitute" is hotly contested (though our reconstruction of "Hello, Kitty," and the complaint in *New York Times v Open AI* show that an AI program can generate substitutional outputs).[154] But *Authors Guild v Google* featured another constraint which reduces its pertinence to AI outputs: the outputs were "transformative," not simply because they repurposed the copied expression to "serve a different function," but because that function continued to target the books from which the snippets drew. The snippets' copied expression identified particular books; "Google's making of a digital copy to provide a search function is a transformative use, which augments public knowledge by making available *information about Plaintiffs' books* without providing the public with a substantial substitute for matter protected by the Plaintiffs' copyright interests in the original works or derivatives of them."[155] The Supreme Court in *AWF* has underscored the importance of a justification rooted in the light the use sheds on the copied work: a use is transformative if "'conjures up' the original work to 'she[d] light' on the work itself, not just the subject of the work."[156] "AWF's use of Goldsmith's photograph does not target the photograph, nor has AWF offered another compelling justification for the use."[157] Similarly, AI outputs may incorporate the source works' expression in a new production; but that output generally will not comment, criticize, shed light on or otherwise be *about* the copied expression. To the extent the outputs incorporate the source works' recognizable expression, the outputs' failure to "comment on, criticize, or otherwise target"[158] that expression weights the first factor against the use.

As a result, it may not be enough to claim that the outputs serve the different purpose of generating new and different text, images, musical compositions, *etc*.; if they "ha[ve] no critical bearing on the substance or style of the original composition,… the claim to fairness in borrowing from another's work diminishes accordingly (if it does not vanish), and other factors, like the extent of its commerciality, loom larger."[159] While, in the absence of a market for training data, the *Authors Guild v Google* decision may support excusing the inputs if the outputs "target" the inputs' expression, the converse would also hold: if the outputs infringe because their reproduction of recognizable expression does not bear on that expression, then the inputs lose the shelter that non infringing outputs might have provided.

---

[152] *Ibid* at 223, 225.

[153] *Ibid* at 222 (emphasis added).

[154] See cases cited, *supra* note 121.

[155] *Authors Guild v Google*, *supra* note 118 at 207 (emphasis added).

[156] *Supra* note 3 at 1281, quoting *Campbell*, *supra* note 27 at 579.

[157] *Supra* note 3 at 1281.

[158] *Ibid* at 1285 n 20.

[159] *Ibid* at 1276, quoting *Campbell*, *supra* note 27 at 580.

3. *Application of these principles to examples of AI-generated images*

To test the above analysis, I requested that the image-generation program DALL-E2 create an image of the Singaporean Merlion eating a durian in a field of durians under a grove of Supertrees in the Gardens by the Bay. DALL-E2 returned the following:

The failure of this output to return an image that resembles either the Merlion (which came out looking more like a dragon), or a durian (which seems more like an avocado), or the Gardens by the Bay Supertrees (here with nonexistent vegetation), may suggest that DALL-E is not simply retrieving and collaging actual images of these items. That in turn may indicate that DALL-E does not "know" what these items look like (i.e. that DALL-E's training data contained insufficient images of these items at the time I formulated my request). But DALL-E did better at representing the Merlion without the durians and the Supertrees:
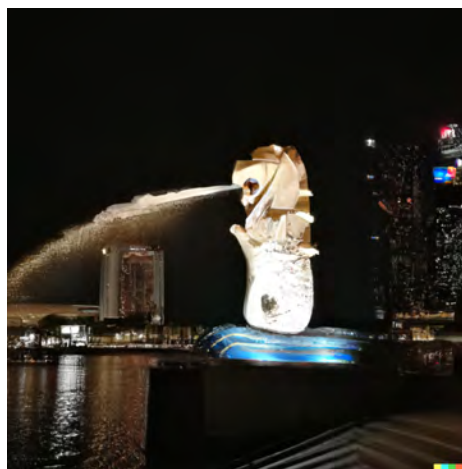
and standalone Supertrees:

and the durians without the Merlion and the Supertrees:
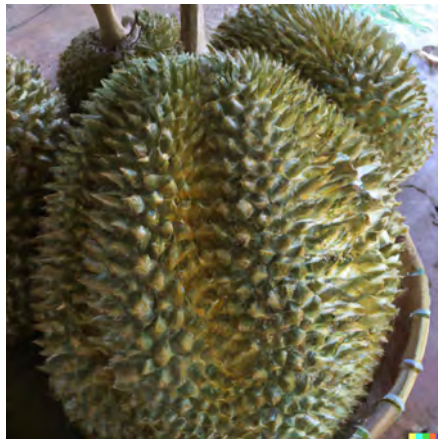
and even the Durians with just the Merlion:

though these last durians look a lot more like lemons than the malodorous fruit.

These results suggest both that some prompts elicit "correct" responses better than others (as was also true for "Hello, Kitty"), and that the visual information requisite to producing outputs that correspond to the prompts does exist in the training data, or can be scraped "live" from the internet. But these results do not tell us whether the images of the Merlion, the durians and the Supertrees were simply reproduced from pre-existing sources stored in DALL-E's database, or whether DALL-E, having "learned from" training data containing images of the Merlion, the durians and the Supertrees, then constructed its own depictions. If DALL-E simply reassembled recognizable excerpts from information in its training data, *AWF* and *Google Books* indicate that the input of entire works may not be excused if the output conveys substantial amounts of the ingested works' expression, without justification, and in competition with the potential licensing of the source works.

By contrast, if the AI program returns an image that falls short of a substantial reconstitution of the source images, a fair use defense to infringement may turn on the extent to which these partial images are either unrecognizable or repurposed in a way that "sheds light" on the copied material.

Finally, suppose the program is referencing the source images as part of a "machine learning" process that enables the program to produce its own images, having "learned" from the source images what a durian looks like. In that event, one might account for the resemblances between the source images and the AI output as the normal consequence of independently produced images that portray the same subject. Copyright abounds in lawsuits that have foundered for plaintiff's failure to prove that the defendant in fact copied her work.[160] In the case of training data, there is upstream copying, but it may not be possible to show that the output in fact incorporated the copied work.

But sometimes it is possible.[161] Suits filed in the UK and in Delaware by Getty Images against Stability AI claim[162] that the text-to-image generator has been reproducing portions of its images, including its watermarks.

---

[160] See *eg*, *Selle v Gibb* 741 F 2d 896 (7th Cir, 1984); *Design Basics, LLC v Lexington Homes, Inc* 858 F 3d 1093 (7th Cir, 2017); *Mag Jewelry Co v Cherokee, Inc* 496 F 3d 108 (1st Cir, 2007); *Repp v Webber* 858 F Supp 1292 (S.D.N.Y. Dist Ct, 1994); *Takeall v Pepsico, Inc* 809 F Supp 19 (D. Md. Dist Ct, 1992).

[161] Matthew Sag, "Copyright Safety for Generative AI" (2023) 61(2) Hous L Rev 101 at 129; Katyanna Quach, "It is possible to extract copies of images used to train generative AI models", *The Register* (6 February 2023). See *eg*, Carlini, *supra* note 117; Gowthami Somepalli *et al*, "Diffusion Art Or Digital Forgery? Investigating Data Replication In Diffusion Models", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022), https://arxivorg/abs/2212.03860; Haim, *supra* note 117.

[162] *Getty Images v Stability AI, Inc,* No 3:23-cv-00201 (filed 3 February 2023) (D. Del.); *Getty Images (US), Inc v Stability AI, Inc*, No 1:23-cv-00135-UNA (filed 3 February 2023) (D. Del.).

An image created by Stable Diffusion showing a recreation of
Getty Images' watermark[163]

Similarly, artist Lauryn Ipsum has found her signature incorporated in images
produced with photo-editing app LensaAI, which combines users' selfies with sup-
posedly stylistic elements inspired by artworks scraped from the Internet.[164]

---

[163] James Vincent, "Getty Images is suing the creators of AI art tool Stable Diffusion for scraping its con-
tent", *The Verge* (17 January 2023).

[164] Morgan Sung, "Lensa, the AI portrait app, has soared in popularity. But many artists question the ethics
of AI art", *NBC News* (6 December 2022).
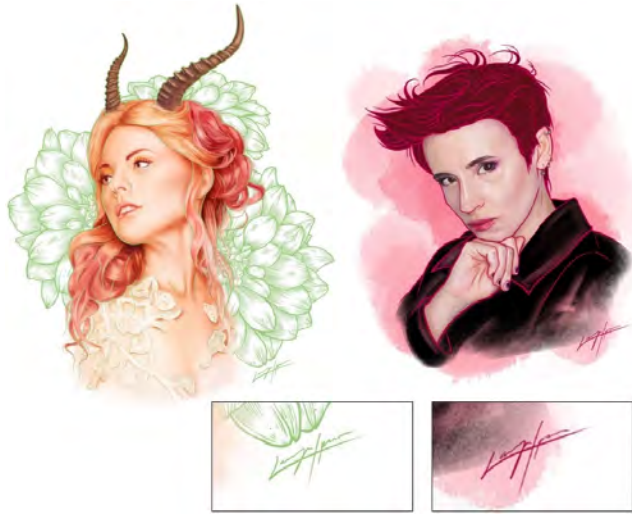
Portraits by illustrator Lauryn Ipsum with insets magnifying the artist's signature[165]
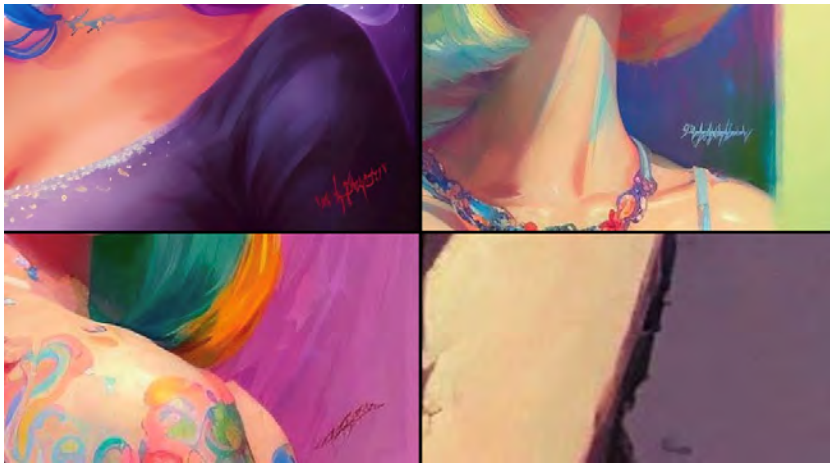
Two Lauryn Ipsum's screenshots of user-generated portraits made on LensaAI,
highlighting the AI's production of signatures[166]

While copyright does not generally protect an artist's style, the line between
style and expression is not always well-marked, and the appearance of an artist's
signature or watermark may belie protestations that the AI system has hewed to the
safe side of the line.

---

[165] To the left, an illustration of "Jenny" based on a photograph by Greg De Stefano, online: Laurtn Ipsum
<https://www.laurynipsum.com/digital/jenny>. To the right, an illustration of "Laurie", online: Lauryn
Ipsum < https://www.laurynipsum.com/digital/laurie>.
[166] Lauryn Ipsum, screenshots of LensaAI portraits, in Shanti Escalante-De Mattei, "Artists Voice Concerns
Over The Signatures In Viral LensaAI Portraits", *ARTnews* (9 December 2022).

Nonetheless, there is a line between style and expression, and emulating a given artist's style has not traditionally been considered copyright infringement.[167] As a result, and assuming the fair use status of the inputs were to turn on the infringement analysis of the outputs, feeding an artist's corpus into an AI system and then asking the system to create a new work in the style of the targeted artist (assuming the work imitates, rather than cuts and pastes from pre-existing works) would not infringe. For example, the "next Rembrandt" is a portrait of a fictitious 17th-century Dutch gentleman, as Rembrandt might have depicted him, had he existed.

The "Next Rembrandt" is a 3D-printed portrait made using deep learning algorithms[168]

To generate the portrait, a team of art historians and of computer scientists created a database of every portrait Rembrandt is known to have painted, and dismembered the images into separate banks of eyes, noses, mouths and facial hair, hats, ruffs and other clothing in order to "teach" the computer to recognize Rembrandt's way of rendering these elements, and then to produce an image that would look like a Rembrandt.[169] Were Rembrandt's works still under copyright, the "Next Rembrandt" output would no more infringe than would a traditional handmade

---

[167] See *eg*, *Hayuk v Starbucks* 157 F Supp 3d 285 (S.D.N.Y. Dist Ct, 2016); *McDonald v West* 138 F Supp 3d 448 (S.D.N.Y. Dist Ct, 2015); *Dean v Cameron* 53 F Supp 3d 641 (S.D.N.Y Dist Ct, 2014); *Kroencke v GMC* 270 F Supp 2d 441 (S.D.N.Y. Dist Ct, 2003) at 444: "Nothing in the Copyright Act of 1976 (which refers to the infringed 'work' in the singular) or in the precedents of this Circuit supports the view that a plaintiff's entire *oeuvre*, or even an aggregated portion of it, may be used as the point of comparison where the works included therein bear little or no relation to one another beyond 'style'); *Judith Ripka Designs, Ltd v Preville* 935 F Supp 237 (S.D.N.Y. Dist Ct, 1996) at 248: "The copyright laws do not protect styles, but only particular original designs" (citing *Peter Pan Fabrics, Inc v Martin Weiner Corp* 274 F 2d 487 (2nd Cir, 1960) at 489).

[168] See https://www.vml.com/work/next-rembrandt.

[169] *Ibid*.

forgery in the style of the Master. Passing the "Next Rembrandt" off as an actual Rembrandt, however, could give rise to a variety of non-copyright claims, for example for false representation.[170] The prospect of illicit uses of a non-infringing output in turn raises questions about the lawfulness of the input, at least when the person or entity compiling the inputs is the same as the person or entity misrepresenting the authorship of the output. At that point, at least for some courts, the fair use doctrine might not excuse the prima facie infringing status of the inputs even if the outputs do not infringe, because fair use implies "good faith and fair dealing,"[171] and that in turn suggests that the conduct should be lawful in general. Because wrongfully attributing authorship and misleading consumers violates a variety of state and federal laws,[172] the fair use claim for the inputs seems very problematic.

The question becomes more complicated if the person misrepresenting the authorship of the output is not the same as the compiler of the inputs. The compiler, whom we shall assume also devised and operates the AI system, has enabled end users to produce images that might be used for illicit purposes, but a compiler lacking actual or "red flag" awareness of the end user's unlawful activities might not be liable for providing the means to engage in those wrongful acts.[173] That conclusion may depend, however, on whether the compiler can deploy upstream means to prevent users from engaging in infringing conduct. One image-generation program, Stable Diffusion Version 2, announced that it had modified the program to make it more difficult to mimic the work of certain artists.[174] In the future, the failure of AI entrepreneurs to forestall at least the most predictable kinds of infringement may amount to a form of actionable wilful blindness to the infringements they enable.

Assume, however, that the end user does not misrepresent her production, and even appends a disclaimer, such as "in the manner of," "homage to" or some other recognized designation of non-provenance. Nonetheless, by generating outputs in the style of a given artist, the end user may, some artists fear, supersede the demand

---

[170] See Lanham Act 15 USC, § 1125 s.43(a).

[171] See sources cited, *supra* note 125. But see empirical studies indicating that bad faith has not had a statistically significant effect on fair use analyses by courts: Barton Beebe, "An Empirical Study of US Copyright Fair Use Opinions Updated, 1978-2019" (2020) 10 NYU J Intell Prop & Ent L 1 at 30; Jiarui Liu, "An Empirical Study of Transformative Use in Copyright Law" [2019] 22(1) Stan Tech L Rev 163 at 168. See also Simon J. Frankel & Matt Kellogg, "Bad Faith and Fair Use" (2012) 60(1) J Copyright Soc'y U.S.A. 1 (arguing that bad faith should not be considered in the fair use analysis); Pierre N. Leval, "Toward a Fair Use Standard" (1990) 103(1) Harv L Rev 1105 at 1126–1128.

[172] At a federal level, see Copyright Law of US, *supra* note 12, §106A; at a state level, see *eg*, Cal. Civ Code § 987(d); Ct. Gen. Stat. § 42-116t(b); La. Rev Stat. Ann. § 51:2154; Mass. Gen. Laws Ann. ch. 231, § 85S(d); Me. Rev Stat. Ann. tit. 27, § 303.03; N.Y. Arts & Cult. Aff. Law § 14.03; N.J. Rev Stat. § 2A:24A-5; 73 Pa. Cons. Stat. Ann. § 2104; R.I. Gen. L. § 5-62-4. For more federal and state protections for moral rights, see US Copyright Office, Authors attribution, and Integrity: Examining Moral Rights in the United States, online: copyright.gov <https://www.copyright.gov/policy/moralrights/>.

[173] See *eg*, *re* common law of secondary liability in copyright, *Gershwin Publishing Corp v Columbia Artists Management, Inc* 443 F 2d 1159 (2nd Cir, 1971) at 1162; *Fonovisa Inc v Cherry Auction, Inc* 76 F 3d 259 (9th Cir, 1996); *BMG Rights Management (US) LLC v Altice USA, Inc,* No 2:22-CV-00471-JRG, 2023 US Dist. LEXIS 84026 (E.D. Tex. Dist Ct, 2023) (summarizing caselaw).

[174] See James Vincent, "Stable Diffusion made copying artists and generating porn harder and users are mad", *The Verge* (24 November 2022).

for the real artist's works.[175] Market-substitutional copying generally does not qualify as a fair use, but in this instance the substitution results from emulating style, rather than incorporating portions of the inputted works into the outputs. Indeed, we have posited that the outputs may not even be prima facie infringing. On the other hand, the AI system would not be able to imitate a given artist's style had it not first copied a great number of that artist's works. The outputs, even if hewing to the "style" side of the "style/expression" divide, moreover may compete with the artist's future prospects: why pay an artist to create a work when one can request an image-generation program to create a similar substitute for free (or for the usage or subscription cost of the service)? But this substitution effect may not be cognizable under factor four of section 107, which inquires into the effect on the market for or value of the copyrighted *work* (i.e., the works copied into the system), not the market for or value of the artist's present and future *oeuvre* in general.

By contrast, under Berne Convention[176] article 9(2), and WIPO Copyright Treaty[177] [WCT] art. 10, the "three-step test" authorizes member states to create exceptions and limitations to the reproduction right; the third step requires that the exception or limitation "not unreasonably prejudice the legitimate interests of the author." An AI output that competes with an artist's *oeuvre* in general or with her future work may not supplant the market for any particular copied work, but that output may indeed "unreasonably prejudice the legitimate interests of the author" in making a living and continuing her creative activities. The relevant interests in the Berne Convention and the WCT focus on the author rather than on particular works.[178] This disparity need not, however, lead to an impasse. The US fair use analysis is not limited to the four factors: courts may take other considerations into account.[179] Whether to avoid putting the US out of compliance with its international obligation to conform national copyright exceptions to the three-step test, or as a matter of purely domestic interpretation of the copyright law, US courts could look

---

[175] See *eg*, Jose Antonio Lanz, "Greg Rutkowski Was Removed From Stable Diffusion, But AI Artists Brought Him Back", *Decrypt* (29 July 2023); Richard Currie, "Musicians threaten to make Oasis 'Live Forever' with AI", *The Register* (21 April 2023); Sarah Andersen, "The Alt-Right Manipulated My Comic. Then A.I. Claimed It", *The New York Times* (31 December 2022); Andy Baio, "Invasive Diffusion: How one unwilling illustrator found herself turned into an AI model", *Waxy* (1 November 2022); Andrew Deck, "AI-generated art sparks furious backlash from Japan's anime community", *Rest of World* (27 October 2022).

[176] Berne Convention (24 July 1971), 828 UNTS 221, (entered into force 19 November 1984, accession by Singapore 21 September 1998) <https://www.wipo.int/wipolex/en/text/283693> [Berne Convention].

[177] WIPO Copyright Treaty (20 December 1996), 36 ILM 65 (1997) (entered into force 6 March 2002, accession by Singapore 17 January 2005) <https://www.wipo.int/wipolex/en/text/295157> [WCT].

[178] By contrast, art.13 of the TRIPS: Agreement on Trade-Related Aspects of Intellectual Property Rights (15 April 1994), Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 UNTS 299, (entered into force 1 January 1995) < https://www.wto.org/english/docs_e/legal_e/27-trips. pdf> formulates the third step with reference to the "rightsholder," thus attenuating the authorship-based interpretation of the Berne Convention, *supra* note 176, and the WCT, *supra* note 177.

[179] See *eg*, *Rogers v Koons* 960 F 2d 301 (2nd Cir, 1992) at 309 (the defendant knowingly infringed in bad faith). See also *MCA, Inc v Wilson* 677 F 2d 180 (2nd Cir, 1981) at 183); *Authors Guild v Google*, *supra* note 118 (the use provides a significant benefit to the public. See also *Perfect 10, Inc v Amazon. com, Inc* 508 F 3d 1166 (9th Cir, 2007) at 1168); *Nunez v Caribbean International News Corp* 235 F 3d 18 (1st Cir, 2000) (the defendant attributes the plaintiff's work to the plaintiff); *Triangle Publications, Inc v Knight-Ridder Newspapers, Inc* 626 F 2d 1171 (5th Cir, 1980) at 1176 (the use is consistent with industry practices).

beyond section 107's lack of an explicit direction to address an exception's impact on the legitimate interests of the *author*. US courts could incorporate an inquiry into that impact as an additional fair use consideration. Finally, even under a solely work-based interpretation of section 107(4), one may observe that the wholesale copying of an artist's works into training data in order to enable stylistically similar outputs jeopardizes not only the artist's future employment or commissions, but also devalues the actual works copied, because the image-generation program can produce outputs that compete with already-created works as well.

In sum, the many factual and legal considerations explored here impede confident prediction of whether the inputs or the outputs would be ruled non infringing. If the lawfulness of the inputs turns on the character of the outputs, one cannot determine either *ex ante*. The same inputs might or might not be fair use depending on different end users' prompts. If an AI system ingests multiple images of apples, including Cézanne's depictions (let's assume Cézanne's works were still under copyright) its training data will enable the system to "know" both what an apple looks like, and what a Cézanne apple looks like. The fair use inquiry may depend on whether the user asks for an apple, or for a Cézanne apple.

IV. Conclusion

Designers of AI systems cannot rely on ex-post assessments of lawful use. A copyright exception, such as set out in Singapore's provisions on Computational Data Analysis,[180] that allows analysis of the inputted contents, but apparently not the generation of outputs for communication to the public, would not suffice. But a specific exception that AI system designers would find adequately broad might strike authors of inputted works as too permissive and livelihood-threatening. In the absence of a specific exception, fair use will not avail AI entrepreneurs if courts focus on the competitive overlap in the assembly of training data: unauthorized inputs impinge on the developing market for licensing works to produce datasets. If courts decline to take into account the market for "information" derived from works of authorship, the fair use status of the inputs will turn on the non-infringing character of the outputs, but fair use will not accomplish the task of immunizing inputs in a way that provides sufficient security to AI entrepreneurs unless they can ensure, upstream, that outputs will not infringe. That in turn means that they must design the system so that no output incorporates qualitatively substantial expression. Like Hathi Trust, the system would need to refrain from output-level copying altogether, or, as Google Books prevented stringing together expressive "snippets," by ensuring that the system renders any "regurgitated" copied material unrecognizable. These guidelines would apply both to outputs the system produces on its own, as well as to outputs requested by end-users. In the latter event, system designers may need to disable features that would allow users to create recognizable copies. In the end, it may prove more cost-effective, as well as more predictable, to license the inputs under terms and conditions that define acceptable outputs.

---

[180] Copyright Act 2021 (2020 Rev Ed) (S'Pore), ss.243, 244.