



**BEFORE THE  
U.S. COPYRIGHT OFFICE**

**Artificial Intelligence and Copyright**

**Docket No. 2023–6**

**REPLY COMMENTS OF THE COPYRIGHT ALLIANCE**

The Copyright Alliance appreciates the opportunity to submit the following reply comments in response to the [notice of inquiry and request for comments](#)<sup>1</sup> (“NOI”) published by the U.S. Copyright Office in the Federal Register on August 30, 2023 (and supplemented by the extension of the reply comment period on November 15, 2023)<sup>2</sup>, regarding the Office’s study of the copyright law and policy issues raised by artificial intelligence (“AI”) systems. These comments are intended to supplement the [comments](#) we filed on with the Office on October 30.<sup>3</sup>

---

<sup>1</sup> Artificial Intelligence and Copyright, 88 Fed. Reg. 59,942 (Aug. 30, 2023).

<sup>2</sup> Artificial Intelligence and Copyright: Extension, 88 Fed. Reg. 78,393 (Nov. 15, 2023).

<sup>3</sup> Copyright Alliance, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-8935>.

With over ten thousand comments filed, it was not possible to review and respond to all the comments that were filed in the time allotted. The fact that we do not respond to a particular comment submitted to the Office during the initial comment period should not be considered as agreement, disagreement, neutrality, or disinterest in the comment. Thus, if, in the process of reviewing all the comments, the Office comes across a particular point that we do not address in our initial or reply comments and is interested in the Copyright Alliance’s view on that issue, we would be pleased to schedule an *ex parte* meeting with the Office to discuss any outstanding issues. Finally, unless stated otherwise, these reply comments focus exclusively on generative AI models.

## I. INGESTION, TRAINING, AND FAIR USE

### 1. *Warhol* Confirms that Transformative Purpose Does Not Control Fair Use Analyses

Many commentors focus their fair use analyses on what they consider to be the transformative purpose of generative AI training and ignore the clear confirmation by the Supreme Court in *Warhol* that transformative purpose does not control a fair use analysis (or even a factor one analysis).<sup>4</sup> In claiming that the training of its Claude model qualifies as fair use, Anthropic says that “[t]his sort of transformative use has been recognized as lawful *in the past* and should continue to be considered lawful in this case.”<sup>5</sup> (emphasis added) This statement may be accurate in that a finding of transformative use in the years before the *Warhol* decision often resulted in an ultimate holding in favor of fair use, but that is far less of a certainty after *Warhol*.<sup>6</sup>

---

<sup>4</sup> See, e.g., Meta Platforms, Inc., Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-9027>; Google, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-9003>; Electric Frontier Foundation, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-8949>; Authors Alliance, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-8976>.

<sup>5</sup> Anthropic PBC, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-9021>.

<sup>6</sup> Jiarui Liu, *An Empirical Study of Transformative Use in Copyright Law*, 22 Stan. Tech. L. Rev. 163 (2019), available at SSRN: <https://ssrn.com/abstract=3330236>.

Moreover, many commentators cite to pre-*Warhol* cases to support their position that a transformative purpose outweighs the commercial nature of a use under factor one.<sup>7</sup> These comments ignore *Warhol's* clarification that transformativeness is merely one subfactor of a factor-one fair use analysis and that commerciality “looms larger” when there is little or no justification for the use.<sup>8</sup>

While many commentators do not address *Warhol's* discussion of justification, at least one claims that there is “overwhelming justification for copying,” which is that “the ability to accurately distill the desired facts (whether about language or images or sounds) requires the ingestion of massive amounts of content that cannot reasonably be individually licensed.”<sup>9</sup> But, as we explain in our initial comments, there are successful AI developers that do not indiscriminately scrape the internet and ingest copyrighted works without permission. Simply because some AI developers want to use everything to train their systems doesn't mean that it is necessary, and their desire to indiscriminately scrape the internet and avoid licensing cannot be a “compelling justification” that would weigh in favor of fair use.

Importantly, many commentators disaggregate the output of generative AI from the training process when discussing the alleged transformative purpose of generative AI. For example, describing the purpose of training, one commentator claims that “copying has a different purpose and message than the creative works being analyzed to generate these observations.”<sup>10</sup> Another commentator says that “[e]xtracting information from training data is the sole purpose and

---

<sup>7</sup> Authors Alliance, *supra* note 4, at 10 (citing to Authors Guild v. Google, 804 F.3d 202 (2d Cir. 2015), to argue that “using copyrighted works to create a tool capable of generating new text and image . . . [is] sufficiently transformative so as to overcome the commercial nature of the entities that use them”); Meta Platforms, Inc., *supra* note 4, at 14 (“Given that wholly distinct purpose, the commercial nature of the use becomes less important.”).

<sup>8</sup> Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 512 (2023).

<sup>9</sup> Meta Platforms, Inc., *supra* note 4, at 7.

<sup>10</sup> Electric Frontier Foundation, *supra* note 4, at 2 (claiming that the purpose of training is to “form observations about patterns arising from a large corpus”).

function of Generative AI model training.”<sup>11</sup> But these comments incorrectly stop their analysis at training and ignore everything in the generative AI process that comes after the initial training—notably the generation of output that serves the same purpose of the ingested works. The *Warhol* decision and many other pre-*Warhol* cases that commentators cite (and we respond to in the next section of these comments) were careful to *consider the ultimate purpose of a use* and not simply end their analysis at an intermediate step—whether that be reverse engineering, shrinking a work into a thumbnail image, researching, or training.

Further demonstrating courts’ considerations of the ultimate or actual purpose of a use, the Second Circuit in *American Geophysical Union v. Texaco, Inc.* rejected over-generalized fair use justifications for widespread infringement, finding that:

The purposes illustrated by the categories listed in section 107 refer primarily to the work of authorship alleged to be a fair use, not to the activity in which the alleged infringer is engaged. Texaco cannot gain fair use insulation for [its employee]’s archival photocopying of articles (or books) simply because such copying is done by a company doing research. It would be equally extravagant for a newspaper to contend that because its business is “news reporting” it may line the shelves of its reporters with photocopies of books on journalism or that schools engaged in “teaching” may supply its faculty members with personal photocopies of books on educational techniques or substantive fields. Whatever benefit copying and reading such books might contribute to the process of “teaching” would not for that reason satisfy the test of a “teaching” purpose.<sup>12</sup>

*Warhol*, *Texaco*, and many of the cases discussed in the section below make clear that, when considering the purpose of the use in a factor-one fair use analysis, the analysis must not merely consider the intermediate step, but rather must take into consideration the ultimate purpose of the use. Thus, the purpose of generative AI cannot be considered in a vacuum of “training.”

---

<sup>11</sup> Meta Platforms, Inc., *supra* note 4, at 12.

<sup>12</sup> *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 924 (2d Cir. 1994).

## 2. Addressing Fair Use Cases Raised in the Initial Comments

Some commentators reference certain court cases for support of their position that AI ingestion of copyrighted works (for training purposes) categorically qualifies as fair use. However, as we explained in our initial comments, determining whether a particular use qualifies for the fair use defense to infringement requires a fact-specific inquiry that is considered on a case-by-case basis. Courts will need to evaluate fair use defenses involving AI systems the same way they evaluate fair use in all contexts: by applying the four factors set forth in section 107 of the Copyright Act to the specific uses at issue.

In our initial comments, we explained why three cases—*Sega v. Accolade*,<sup>13</sup> *Sony v. Connectix*,<sup>14</sup> and *Authors Guild v. Google*<sup>15</sup> (the Google Books case)—fail to support a categorical fair use exception for AI ingestion. A handful of other court decisions were referenced in some of the initial comments that we did not address in our initial comments. Below we discuss those cases and explain why they likewise fail to support a categorical fair use exception for AI ingestion, and in fact demonstrate that current copyright law is fit for purpose to analyze whether the fair use exception applies, which should be decided on a case-by-case basis.

***Kelly v. Arriba Soft***.<sup>16</sup> In this Ninth Circuit case, Arriba Soft was sued for copyright infringement for scraping plaintiff’s photographs from the internet and then displaying smaller, lower-resolution “thumbnail” copies of the photographs on the search results page of its visual search engine. The court held that Arriba Soft’s reproduction of plaintiff’s photos as thumbnail images qualified as fair use because the thumbnail images served an entirely different purpose than the original images. Specifically, the court found that plaintiff’s photographs were artistic works that “were intended to inform and to engage the viewer in an aesthetic experience,” in contrast to Arriba’s use which the court found was “unrelated to any aesthetic purpose” and that the search

---

<sup>13</sup> *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992).

<sup>14</sup> *Sony Comput. Ent., Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000).

<sup>15</sup> *Authors Guild v. Google (Google Books)*, 804 F.3d 202 (2d Cir. 2015).

<sup>16</sup> *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003).

engine instead “functions as a tool to help index and improve access to images on the internet and their related web sites.”<sup>17</sup> Had the purposes of plaintiff’s images and defendant’s images been the same, the result in this case would have been very different.<sup>18</sup> Thus, contrary to those commentators who reference this case to try to establish that this decision stands for the proposition that AI ingestion is categorically fair use, the decision in this case actually stands for the opposite proposition because, in the context of AI, both the purpose of the works of visual art that are ingested by the AI and the purpose of the visual art generated by an AI serve the exact same purpose—“to inform and to engage the viewer in an aesthetic experience.” Because both the ingested work and AI-generated output serve the same purpose, that “seriously weakens” the fair use claim.<sup>19</sup>

The Court also found in favor of fair use because “[t]he thumbnails do not stifle artistic creativity because they are not used for illustrative or artistic purposes and therefore *do not supplant the need for the originals.*”<sup>20</sup> (emphasis added) This is very different from AI ingestion. Unlike the thumbnail images at issue in *Arriba*, when visual art is generated by an AI model it is certainly possible that such AI-generated output would “supplant the need for the original [i.e., the ingested work].”<sup>21</sup>

Many of the commentators disaggregate the purpose of AI systems in their fair use analyses by focusing almost exclusively on the ingestion/training process without acknowledging what the ingestion/training is in service of. However, it is essential to recognize that creating the thumbnails, like copying for generative AI training, involves a process that results in an output. When the court analyzed the fair use defense in *Arriba*, it did not simply analyze the purpose of the intermediate copying. Rather, it considered the purpose of the output and what effect it might

---

<sup>17</sup> *Id.* at 818.

<sup>18</sup> *Id.* at 819 (“[W]here the use is for the same intrinsic as [the copyright holder’s] . . . such use seriously weakens a claimed fair use.”) (citing *Worldwide Church of God v. Philadelphia Church of God, Inc.*, 227 F.3d 1110, 1117 (9th Cir. 2000)).

<sup>19</sup> *Id.*

<sup>20</sup> *Id.* at 820.

<sup>21</sup> *Id.*

have on the need for the original, and creativity in general. In *Arriba*, Arriba Soft copied the image, shrank the image (into a thumbnail), and displayed the thumbnail image. The court considered the ultimate purpose of displaying the images on the search page, not the purpose of the shrinking process itself. In the context of an image generator, an AI company copies the image (typically from the internet) and trains the AI on the image for the purpose of enabling the AI system to generate and display an image. Consistent with the approach taken by the Ninth Circuit in the *Arriba* case (and numerous other cases), a court’s focus should be on the ultimate purpose of generating and displaying an image by the AI model, not solely on the purpose of the intermediate step of training the model.

***Perfect 10 v. Amazon***:<sup>22</sup> In this Ninth Circuit case, plaintiff Perfect 10 sued Google and Amazon for, among other things, infringing its copyrighted images by displaying smaller, lower-resolution “thumbnail” copies of the images on the search results page of Google’s image search. Similar to *Arriba*, the court in *Perfect 10* concluded that the use of thumbnail versions of plaintiff’s images in the search engine qualified as fair use because, “[a]lthough an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing the user to a source of information.”<sup>23</sup> This statement clearly distinguishes the facts in both *Perfect 10* and *Arriba* from generative AI fair use cases that might arise because the purpose of AI copying has nothing to do with pointing a user to a source of information.

***AV ex rel. Vanderhuy v. iParadigms***:<sup>24</sup> In this Fourth Circuit case, iParadigms, a company that offers a plagiarism detection service to enable schools to monitor for plagiarism by digitally comparing students’ works, was sued for copyright infringement by high school students whose papers had been submitted to iParadigm by schools and archived by iParadigms for continued use in these digital comparisons. The court held that iParadigm’s archiving of student work for the purpose of detecting plagiarism constituted fair use because “iParadigms’ use of plaintiffs

---

<sup>22</sup> *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9<sup>th</sup> Cir. 2007).

<sup>23</sup> *Id.* at 1165.

<sup>24</sup> *AV ex rel. Vanderhuy v. iParadigms*, 562 F.3d 630 (4<sup>th</sup> Cir. 2009).

works had an entirely different function and purpose than the original works,” and “was not related to the creative core of the works.”<sup>25</sup> Unlike *iParadigms*, and as noted in *Arriba* above, generative AI is ingesting copyrighted works to generate works that will almost certainly serve a similar “function and purpose as the original [ingested] works.”

Also, under the fourth fair use factor, the *iParadigms* court found that there was no market whatsoever for the high school papers because creating a market to sell high school papers would be considered to be “dishonest” and “cheating” and would “encourage plagiarism” amongst students and thus, “any harm [to the market] is not the kind protected against by copyright law.”<sup>26</sup> Given these extremely unique facts in the *iParadigms* case, the fourth factor analysis in *iParadigms* is completely inapplicable to AI. Moreover, as explained in more detail in our initial comments, there is a burgeoning market for copyrighted works for ingestion and, unlike in *iParadigms*, there is nothing untoward about that market.

As a result, while the *iParadigms* case is somewhat instructive because it is one of numerous cases that teaches courts to consider the *ultimate* purpose of the use, the decision itself is inapplicable to any AI fair use analysis. What is clear is that it certainly does not stand for the proposition that AI ingestion/training is categorically fair use because the facts at issue in the cases and the purposes are completely different.

***Authors Guild v. HathiTrust***:<sup>27</sup> Google, in conjunction with several universities, created HathiTrust to administer a digital repository of more than ten million works called the HathiTrust Digital Library (HDL) that is used, among other things: (i) for full-text searching by the public and (ii) to allow library patrons with print disabilities to have access to full texts of works. The Second Circuit held that the creation of a full-text searchable database qualified as fair use, noting that, under the first fair use factor, “the result of a word search is different in purpose,

---

<sup>25</sup> *Id.* at 639, 641.

<sup>26</sup> *Id.* at 634-44.

<sup>27</sup> *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014).



character, expression, meaning, and message from the page (and the book) from which it is drawn”<sup>28</sup> and, under the third factor, that copying was reasonably necessary to facilitate this search service.<sup>29</sup> The court also held that HDL’s copying for the purposes of providing access to the print-disabled also qualified as fair use.<sup>30</sup> The court found that such use was not a transformative use under the first factor because “the underlying purpose of the HDL’s use is the same as the author’s original purpose.”<sup>31</sup> Nevertheless, the court concluded that providing such access was a valid purpose under the first fair use factor because “making a copy of a copyright works for the convenience of a blind person is expressly identified by the House Committee Report as an example of fair use....”<sup>32</sup> Finally, the court found that the fourth factor weighed in favor of fair use because the market for books accessible to the handicapped is so insignificant that “it is common practice in the publishing industry for authors to forego royalties that are generated through the sale of books manufactured in specialized formats for the blind...”<sup>33</sup>

The court’s analysis in the HathiTrust case seems wholly inapplicable to AI ingestion of copyrighted works because: (i) the purpose of the HDL searchable database—to point a user to information about a work while not competing with the work—is entirely different than the purpose of AI models, which is to generate images, text, or music in a way that may compete with the ingested works; (ii) there is no legislative history suggesting that AI ingestion should be considered to be fair use; and (iii) the market for specialized formats is completely different than the AI ingestion market in that it serves a much smaller-scale, specific community. Thus, any market harm that might occur from HDL is negligible in comparison to that which may result from the unauthorized use of massive amounts of copyrighted works from AI ingestion.

---

<sup>28</sup> *Id.* at 97.

<sup>29</sup> *Id.* at 98.

<sup>30</sup> *Id.* at 100–03.

<sup>31</sup> *Id.* at 101.

<sup>32</sup> *Id.* at 102.

<sup>33</sup> *Id.* at 103.

*Field v. Google*:<sup>34</sup> In this case before the district court in Nevada, plaintiff Field sued Google for copyright infringement for Google’s making of cached copies of his works by scraping plaintiff’s website and making those copies available through Google’s search engine. The court held that Google’s making of cached copies was fair use in part because caching served the purposes of “enabl[ing] users to access content when the original page is inaccessible,” “allow[ing] Internet users to detect changes that have been made to a particular Web page over time,” and “allow[ing] users to understand why a page was responsive to their original query.”<sup>35</sup> The fact that Google employed numerous safeguards to prevent harm to Field and others whose materials are made available through the search engine’s cache functionality—such as “mak[ing] it clear that [Google] does not intend a ‘Cached’ link of a page to substitute for a visit the original page” and “ensur[ing] that any site owner can disable the cache functionality for any of the pages on its site”—also played a role in the fair use analysis.<sup>36</sup> Another very important consideration playing into the fair use analysis is the fact that Field opted in to search engine copying and caching his works by proactively “add[ing] the robots.txt file to his site to ensure that all search engines would include his Web site in the search listings.”<sup>37</sup>

Lastly, the court noted that under the fourth fair use factor, “there was no evidence of any market for Field’s works” because “Field makes his works available to the public for free” and “never received any compensation from selling or licensing them.”<sup>38</sup> Importantly, the court also found that the fourth fair use factor weighed in favor of Google because “there is no evidence before the Court of any market for licensing search engines the right to allow access to Web pages through ‘Cached’ links, or evidence that one is likely to develop.”<sup>39</sup>

---

<sup>34</sup> *Field v. Google Inc.*, 412 F. Supp. 2d. 1106 (D. Nev. 2006).

<sup>35</sup> *Id.* at 1118-1119.

<sup>36</sup> *Id.* at 1119.

<sup>37</sup> *Id.* at 1120. “A robots.txt file tells search engine crawlers which URLs the crawler can access on your site. This is used mainly to avoid overloading your site with requests; it is not a mechanism for keeping a web page out of Google. To keep a web page out of Google, [block indexing with noindex](#) or password-protect the page.” *Introduction to robots.txt*, Google, <https://developers.google.com/search/docs/crawling-indexing/robots/intro> (last updated Nov. 28, 2023).

<sup>38</sup> *Field v. Google Inc.*, 412 F. Supp. 2d. at 1121.

<sup>39</sup> *Id.* at 1122.

In short, the facts in *Field* could not be more different than the facts in a typical AI ingestion case. Unlike in *Field*:

- The purpose of AI copying is not for any of the functional justifications associated with caching, but instead the purpose is to generate work of a similar nature to the ingested work that may serve as a substitute for the original and adversely impact the marketplace.
- Many AI companies do not implement safeguards to limit or prevent harm to the rights holders of ingested works.
- The copyright owners whose works are being scraped from the internet for ingestion into an AI are not opting in because there is presently no mechanism for doing so, as robots.txt is designed for search engine scraping and there is no AI-specific mechanism to allow for scraping for AI purposes.
- There is an emerging market for licensing copyrighted works to be ingested for training purposes.

Furthermore, it should also be noted that this case is representative of the first factor having an undue influence over the entire fair use analysis—an approach that the Supreme Court in *Warhol* rejected.

When considering the cases referenced above and any other fair use cases in which transformative use plays a significant role in the court's fair use decision, the Copyright Office should take into account whether the cases were decided between the *Campbell* decision and the *Warhol* decision, because during that interim period a determination that a use was transformative was often given undue breadth and the impact of a use being transformative was

given more weight by courts than it will post-*Warhol*.<sup>40</sup> As a result, it is now questionable whether a pre-*Warhol* fair use case in which the determination turned on a finding of transformativeness remains good law.

### 3. AI Ingestion—Like All Other Uses—Is Not Categorically Fair Use

Several commentators claim that AI ingestion is categorically fair use because it is merely an “intermediate step” in “non-expressive” copying<sup>41</sup> or it constitutes “computational analysis.”<sup>42</sup> These claims are incorrect for several reasons.

Professors Samuelson, Sprigman and Sag explain in their comments that “non-expressive use” is one in which “technical acts of copying [] do not communicate an author’s original expression to a new audience.”<sup>43</sup> They say that courts have “consistently held” these “non-expressive” uses qualify as fair use. In support they cite to the same cases we have discussed numerous times in these comments and our initial comments—*Authors Guild v. Google*, *Sega*, *Connectix*, *iParadigm*, *HathiTrust*, *Field*, *Perfect 10* and *Arriba*. For reasons stated above and in our initial comments, these cases do not stand for the broad proposition that there is a so-called “non-expressive use” category of fair use. There are no uses that always categorically qualify as fair

---

<sup>40</sup> In the interim period, a determination that a use was transformative almost always meant that a court concluded it is also fair use. A 2011 study found that of all the fair use cases decided in 2006 through 2010, the defendant won 100% of the time when the court found the subject use to be transformative. Neil Weinstock Netanel, *Making Sense of Fair Use*, 15 Lewis & Clark L. Rev. 715, 754–55 (2011). Between 2010 and 2023, the Copyright Alliance is aware of only one circuit court decision holding that a use was transformative but not fair use. *See* Fox News Network, LLC v. TVEyes, Inc., 883 F.3d 169, 180–81 (2d Cir. 2018); *see also* Liu, *supra* note 6, at 185–86 (finding that transformative use was the only statistically significant subfactor driving the first factor determination).

<sup>41</sup> *See, e.g.*, Anthropic PBC, *supra* note 5; Electric Frontier Foundation, *supra* note 4; Meta Platforms, Inc., *supra* note 4; Pamela Samuelson, Christopher Sprigman, & Matthew Sag, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), *available at* <https://www.regulations.gov/comment/COLC-2023-0006-8854>.

<sup>42</sup> *See, e.g.*, Google, *supra* note 4; Computer & Communications Industry Association (CCIA), Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), *available at* <https://www.regulations.gov/comment/COLC-2023-0006-8740>; The Software Alliance, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), *available at* <https://www.regulations.gov/comment/COLC-2023-0006-8688>.

<sup>43</sup> Samuelson, Sprigman, & Sag, *supra* note 41, at 11.

use. More specifically, neither the courts nor Congress have ever espoused a broad category of fair use called “non-expressive use.” It’s not in the preamble or anywhere in section 107 and it is not found in any case law. In fact, such a categorical approach is antithetical to the fair use doctrine. Each fair use case must be considered and decided based on its own set of facts.

Most of the cases referenced in support of the proposition that AI ingestion is categorically fair use because the use at issue in the case was found to be transformative can be placed into one of two buckets: (i) search engine cases (e.g., *Authors Guild*, *HathiTrust*, *Arriba*, *Perfect 10* and *Field*), and (ii) reverse engineering for interoperability cases (*Sega* and *Connectix*).<sup>44</sup> There is no broad “non-expressive use” bucket.

Despite the fact that some cases involving use for search engines or reverse engineering purposes involved facts that resulted in a finding of fair use, like all fair use analyses, these uses are decided on a case-by-case basis, taking into consideration the specific facts at issue. As such, there are also cases in which reverse engineering and search engine use do not qualify as fair use. For example, in *Associated Press v. Meltwater*<sup>45</sup> the Southern District of New York held that Meltwater was more than just a search engine and its use was not fair use. In so holding, the court clearly stated that:

[E]ven if it were a search engine it would still be necessary to examine whether Meltwater had acted to violate the Copyright Act... In other words, using the mechanics of search engines to scrape material from the Internet and provide it to consumers in response to their search requests does not immunize a defendant from the standards of conduct imposed by law through the Copyright Act, including the statutory embodiment of the fair use defense.<sup>46</sup>

---

<sup>44</sup> We did not include *iParadigms* here because the factual scenario is so unique—archiving copies for certain schools to use to detect student cheating—that there is only one case that falls into that category. Nor do we include the making of specialized format versions of copyright books, as takes place in *HathiTrust*, because that use was found not to be transformative.

<sup>45</sup> *Associated Press v. Meltwater U.S. Holdings, Inc.*, 931 F. Supp. 2d 537 (S.D.N.Y. 2013).

<sup>46</sup> *Id.* at 556.

In another search engine case, *Fox News Network v. TVEyes*, the Second Circuit explained that the case “shares features with [the] decision in *Authors Guild v. Google*” and noted in its decision that it had “cautioned that the [*Authors Guild*] case ‘tested the boundaries of fair use.’”<sup>47</sup> In *TVEyes*, the Second Circuit concluded that, despite operating a search engine and sharing similarities with the *Authors Guild*’s case, TVEyes “exceeded those bounds” because TVEyes “failed to show that the product it offers to its clients can be justified as a fair use.”<sup>48</sup> The TVEyes case demonstrates the importance of considering the facts of each case independently and the justification for the use, and not classifying something as fair use simply because it either (i) falls into a category in which other cases have found fair use or (ii) it shares factual similarities with other cases—such as the *Authors Guild* case, which by the Second Circuit’s own admission “test[ed] the boundaries of fair use.”<sup>49</sup>

“Intermediate copying” and “reverse engineering” are likewise not uses that categorically qualify as fair use. For example, in *Atari v. Nintendo*<sup>50</sup> (as discussed in more detail in our initial comments), the Federal Circuit held that Atari’s intermediate copying of copyright protected source and object code for reverse engineering purposes did not qualify as fair use. There are numerous other examples of cases that did not result in a finding of fair use. The point is, as numerous courts have stated throughout time, courts must consider the facts of each case independently. And thus, conclusory statements that “intermediate copying,”<sup>51</sup> “computational

---

<sup>47</sup> *Fox News Network, LLC v. TVEyes, Inc.*, 883 F.3d 169, 174 (2d Cir. 2018).

<sup>48</sup> *Id.*

<sup>49</sup> *Authors Guild v. Google (Google Books)*, 804 F.3d 202, 206 (2d Cir. 2015).

<sup>50</sup> *Atari Games Corp. v. Nintendo of Am. Inc.*, 975 F.2d 832 (Fed. Cir. 1992).

<sup>51</sup> *See Walt Disney Productions v. Filmation Assoc.*, 628 F. Supp. 871 (1986) (holding that the Copyright “Act prohibits the creation of copies, even if the creator considers those copies mere interim steps toward some final goal”); *see also Walker v. University Books, Inc.*, 602 F.2d 859 (9th Cir. 1979) (“[T]he fact that an allegedly infringing copy of a protected work may itself be only an inchoate representation of some final product to be marketed commercially does not in itself negate the possibility of infringement.”).

analysis,”<sup>52</sup> “non-exploitive use,” “reverse engineering,” “non-expressive use” and “AI ingestion” categorically qualify as fair use are fatally flawed.

Second, the cases referenced by the commentators held for fair use for reasons that are inapplicable to generative AI. The search engine cases referenced by commentators found for fair use because the search engines pointed back to the underlying material, rather than creating something competitive. Generative AI does not point back to ingested material; it creates something that may compete with the ingested works. The reverse engineering cases found for fair use because the copyrighted works copied (i) were computer programs, which the court afforded a “lower degree of protection than more traditional literary works”<sup>53</sup>; (ii) were not available through licenses; (iii) were for interoperability purposes; and perhaps, most significantly, (iv) were to prevent anti-competitive behavior. AI ingestion shares none of these features. In fact, as discussed in more detail in the comments submitted by the FTC, by potentially harming “creators’ ability to compete,” AI ingestion may have the exact opposite competitive effect of *Sega* and *Connectix*.<sup>54</sup>

Third, even if such a categorical fair use approach were supported by the cases, it wouldn’t apply to generative AI ingestion because AI use is not a so-called “non-expressive use” (which itself is not a categorical exception.) Unlike each of the search engine and reverse engineering cases, generative AI systems ingest and use works for expressive purposes. There is nothing “non-expressive” about them. In fact, the difference between generative AI and analytical AI is that

---

<sup>52</sup> We do not specifically address “computational analysis” because, although CCIA claims that “Judge Leval’s opinion in *Google* provides the clearest analysis of why the creation of datasets for computational analysis, and their subsequent uses in AI training, are fair uses” the term “computational analysis” is nowhere in Leval’s opinion.

<sup>53</sup> *Sega* at 977.

<sup>54</sup> Federal Trade Commission, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), at 5, *available at* <https://www.regulations.gov/comment/COLC-2023-0006-8630> (“As the courts apply the doctrine of fair use to the training and use of AI, the evolution of the doctrine could influence the competitive dynamics of the markets for AI tools and for products with which the outputs of those tools may compete. Conduct that may violate the copyright laws—such as training an AI tool on protected expression without the creator’s consent or selling output generated from such an AI tool, including by mimicking the creator’s writing style, vocal or instrumental performance, or likeness—may also constitute an unfair method of competition or an unfair or deceptive practice, especially when the copyright violation deceives consumers, exploits a creator’s reputation or *diminishes the value of her existing or future works*, reveals private information, or otherwise causes substantial injury to consumers.”) (emphasis added).

generative AI is intended to produce *expressive* output—like literary works, music, art, audiovisual works—whereas analytical AI does not. If those works were not expressive, they wouldn’t be listed as protectable subject matter under section 102 of the Copyright Act.

#### 4. Various Misstatements about Fair Use

Throughout the comments there were various misstatements made about fair use. We address the most egregious of these statements in this section.

Open AI states that “when technical realities require that copyrighted works be reproduced in order to extract and learn from these unprotectable aspects of a work, courts have routinely found those reproductions to be permissible under the fair use doctrine.”<sup>55</sup> In support of this statement they reference three cases: *Sega*,<sup>56</sup> *Connectix*,<sup>57</sup> and *Assessment Technologies of WI, LLC v. WIREdata*.<sup>58</sup> Despite OpenAI’s broad characterization of these cases in an attempt to make them appear as if they establish some general rule for fair use, as we explain in our initial comments, *Sega* and *Connectix* are both reverse engineering/interoperability cases that have little bearing on the fair use analysis of AI ingestion. We discussed a similar reverse engineering case, *Atari v. Nintendo*, where fair use was not found.<sup>59</sup> In short, there is no general rule that makes all reverse engineering fair use, and there certainly is no general rule that applies when “technical realities require that copyrighted works be reproduced in order to extract and learn from these unprotectable aspects of a work.”<sup>60</sup>

---

<sup>55</sup> OpenAI, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), at 12, available at <https://www.regulations.gov/comment/COLC-2023-0006-8906>.

<sup>56</sup> *Sega*, 977 F.2d 1510.

<sup>57</sup> *Sony Comput. Ent., Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000).

<sup>58</sup> *Assessment Technologies of WI, LLC v. WIREdata*, 350 F. 3d 640, 644-45 (7th Cir. 2003).

<sup>59</sup> *See Atari Games Corp. v. Nintendo of Am. Inc.*, 975 F.2d 832 (Fed. Cir. 1992).

<sup>60</sup> Open AI, *supra* note 55.



Further, as we note above when discussing how courts analyze the purpose of a use, in the reverse engineering cases, the courts considered the ultimate purpose of the works created from the process of reverse engineering. The courts did not end their analysis at the reverse engineering step, and nor should courts analyzing fair use claims related to generative AI stop at the “training” step. The third case referenced here, *Assessment Technologies of WI, LLC v. WIREDATA*, is inapplicable to any discussion of fair use as it was not raised as a defense and the court did not opine on fair use in that case.<sup>61</sup>

In comments filed by Professors Samuelson, Sprigman and Sag, they say that “the Copyright Office itself has recognized the fair use status of TDM research.”<sup>62</sup> To support this they note that, in evaluating the proposed DMCA § 1201 exemption to circumvent technological protection measures on DVDs and eBooks for the purpose of conducting text and data mining (TDM), the Copyright Office said: “Balancing the four fair use factors, with the limitations discussed, the Register concludes that the proposed use is likely to be a fair use.”<sup>63</sup> It’s important that the phrase “with the limitations discussed” not be overlooked—especially when considering TDM in the context of generative AI ingestion. The Copyright Office did not broadly state that TDM was “likely to be fair use.” They conditioned their statement on several “limitations” that were delineated in the Report.<sup>64</sup> The most significant of these limitations, for AI purposes is:

- “[T]hat circumvention be permitted only on copies of the copyrighted works that were lawfully acquired and that the institution owns or for which it has a non-time-limited license. Circumvention is not permitted to access copies that the institution has rented or borrowed.”<sup>65</sup>

---

<sup>61</sup> *Assessment Techs.*, 350 F.3d at 644–45 (involving a collection of uncopyrightable data that the defendant copied from the plaintiff’s compilation).

<sup>62</sup> Samuelson, Sprigman, & Sag, *supra* note 41, at 7.

<sup>63</sup> *Id.*

<sup>64</sup> See U.S. COPYRIGHT OFF., SECTION 1201 RULEMAKING: EIGHTH TRIENNIAL PROCEEDING, RECOMMENDATION OF THE REGISTER OF COPYRIGHTS, 121–22 (2021), [https://cdn.loc.gov/copyright/1201/2021/2021\\_Section\\_1201\\_Registers\\_Recommendation.pdf](https://cdn.loc.gov/copyright/1201/2021/2021_Section_1201_Registers_Recommendation.pdf).

<sup>65</sup> *Id.* at 121.

- “[I]t [is] appropriate to add a limitation that the person undertaking the circumvention view or listen to the contents of the copyrighted works in the corpus solely for the purpose of verification of the research findings, not for their expressive purposes.”<sup>66</sup>
- “[T]he institution storing or hosting a corpus of copyrighted works [is] required to use “effective security measures” to secure the corpus from unauthorized access, distribution, or download.”<sup>67</sup>

Significantly, AI ingestion typically fails to meet each of these limitations. Thus, the stand-alone statement that “the Copyright Office itself has recognized the fair use status of TDM research” is misleading without also explaining the limitations to such so-called status.<sup>68</sup> The Copyright Office has not declared that all TDM is fair use, as this statement makes it appear. If anything, one possible takeaway from the 1201 rulemaking is that TDM use is likely not fair use unless the limitations set forth in the rulemaking are each met.

## **5. The Impact of AI Ingestion of Pirated Works Scraped from Rogue Websites**

As we said in our initial comments, if AI developers know or should have known they are ingesting works that have been made available illegally, a fair use defense would be much less likely to succeed. Although the case law related to considerations of good and bad faith in a fair use analysis do not provide much clarity, the fact that an alleged infringer knew or should have known they were copying from works that have been made available illegally is certainly taken into consideration. Unless there is some countervailing factual justification (such as when there is no other way to obtain the work),<sup>69</sup> copying from infringing works is typically weighed against

---

<sup>66</sup> *Id.* at 121-22.

<sup>67</sup> *Id.* at 122.

<sup>68</sup> Samuelson, Sprigman, & Sag, *supra* note 41, at 7.

<sup>69</sup> See e.g., *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1527-28 (9th Cir. 1992).

a particular use being deemed fair use. The analysis should be no different for a fair use analysis for AI ingestion.

The comments by Professors Samuelson, Sprigman, and Sag suggest otherwise, saying that “there are strong arguments to be made that copying from an infringing source may still be fair use.”<sup>70</sup> We do not disagree that “copying from an infringing source *may* [] be a fair use,”<sup>71</sup> because, as we have said numerous times, fair use is very fact dependent and there are no categorical distinctions in a fair use analysis. However, we take issue that to support their “strong arguments” position<sup>72</sup> they reference a paper by Professor Michael Carrol who in turns misleadingly relies on three cases (*Kepner-Tregoe, Inc. v. Leadership Software, Inc.*,<sup>73</sup> *Liu v. Price Waterhouse LLP*,<sup>74</sup> and *Real View, LLC v. 20-20 Techs., Inc.*<sup>75</sup>) when in fact *none of these cases even considered fair use* or the so-called “fruit of the poisonous tree” doctrine referenced in the Samuelson et. al. comments.<sup>76</sup> If there are “strong arguments” to be made here, as Samuelson et. al. say, then one would think that they could at least find cases that reference fair use. Thus, while copying from an infringing source may be fair use, when it is justified, there is no such justification in the generative AI context because there are many alternatives to copying works for ingestion that do not involve scraping infringing works from pirate (or non-pirate) websites. It’s also important to understand that the cases that have addressed copying from works that were made available illegally deal with situations in which a single work was infringed. These cases do not address the use of potentially massive amounts of pirated works are used without authorization—as is the case with AI ingestion.

---

<sup>70</sup> Samuelson, Sprigman, & Sag, *supra* note 41, at 23.

<sup>71</sup> *Id.*

<sup>72</sup> *Id.* at 25.

<sup>73</sup> *Kepner-Tregoe, Inc. v. Leadership Software, Inc.*, 12 F.3d 527, 538 (5th Cir. 1994).

<sup>74</sup> *Liu v. Price Waterhouse LLP*, No. 97 CV 3093, 2000 WL 1644585, at \*2 (N.D. Ill. Oct. 30, 2000).

<sup>75</sup> *Real View, LLC v. 20-20 Techs., Inc.*, 811 F. Supp. 2d 553, 561 (D. Mass. 2011).

<sup>76</sup> Samuelson, Sprigman, & Sag, *supra* note 41, at 25.

## 6. Application of the Third Fair Use Factor

In response to question 8.4, with regard to the third fair use factor, several commentators argue that the “amount and substantiality” of the use of each individual piece of training data is negligible because each piece has a negligible influence on the model’s store of statistical information.<sup>77</sup> This is incorrect for two reasons. First, the amount of individual pieces of training data is irrelevant to the ingestion fair use analysis because entire copies of works are being ingested. Second, and most importantly, the third fair use factor is concerned with the “the amount and substantiality of the portion used *in relation to the copyrighted work as a whole*” (emphasis added).<sup>78</sup> Thus, the amount used is considered in relation to the copyrighted work, not in relation to the alleged infringing work or in relation to the number of other works used without authorization.<sup>79</sup> As a result, whether the amount used constitutes a small or large percentage of the alleged infringing work or has a large or small “influence” on the infringing work is irrelevant to the third fair use factor.

## 7. Generative AI Models Copy Protected Expression

A recurring claim made in many initial comments is that, rather than copying protected expression from ingested works, AI models merely copy unprotectable data or information about the “relationships” and “patterns” amongst expressive elements.<sup>80</sup> For example, in Anthropic’s comments, the AI developer claims that its Claude model is “trained by deriving facts, patterns,

---

<sup>77</sup> See e.g., Library Copyright Alliance, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright, at 1 (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-8452>; Meta Platforms, Inc., *supra* note 4 at 15; and New Media Rights Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright, at 17 (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-9123>.

<sup>78</sup> 17 U.S.C. § 107.

<sup>79</sup> *Id.*; see also Harper & Row, Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 565 (1985) (“As the statutory language indicates, a taking may not be excused merely because it is insubstantial with respect to the infringing work . . . Conversely, the fact that a substantial portion of the infringing work was copied verbatim is evidence of the qualitative value of the copied material, both to the originator and to the plagiarist who seeks to profit from marketing someone else’s copyrighted expression.”).

<sup>80</sup> See, e.g., Meta Platforms, Inc., *supra* note 4; Google, *supra* note 4; Electric Frontier Foundation, *supra* note 4; Anthropic PBC, *supra* note 5.

relationships, concepts, and other uncopyrightable information.”<sup>81</sup> While we explain in our initial comments why that is incorrect, it is a critical point that bears repeating.

AI developers seek out expressive works for ingestion specifically because of the value they provide, and that value is directly attributable to the creative, expressive elements that qualify the works for copyright protection in the first place.<sup>82</sup> The relationships between the words an author chooses for a literary work, the chords a songwriter chooses for a song, and the many expressive elements that go into a work of visual art are all the exact type of expression that is protected by copyright. Simply because these copyrighted works can be processed into an AI model does not mean that the work should lose its protection or that a copyright owner should lose their right to control or be compensated for use of that work. Moreover, it has been shown that AI models like Anthropic’s Claude reproduce verbatim text from ingested works,<sup>83</sup> further demonstrating that the protected expressive elements of a copyrighted work are being copied wholesale during the ingestion process.

Another important and related point is that even if, as some commentators argue, generative AI models engage in “intermediate” copying, such copying can still be infringing. In *Disney v VidAngel, Inc.*, “VidAngel claim[ed] that their copies of Plaintiffs’ works are only ‘intermediate’ copies and not ‘copies’ as defined by the Copyright Act.”<sup>84</sup> The court explained VidAngel’s process of copying as follows: “VidAngel creates ‘intermediate’ files by tag[ging] the files for over 80 types of content, and break[ing] them into approximately 1,300 fragments that contain no more than 10 seconds of content, then encrypt[ing] those fragments, and stor[ing] them in a

---

<sup>81</sup> Anthropic PBC, *supra* note 5, at 2.

<sup>82</sup> See generally NEWS MEDIA ALLIANCE, 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 8–12 (2023), <http://www.newsmediaalliance.org/wp-content/uploads/2023/10/AI-White-Paper-with-Technical-Analysis.pdf>.

<sup>83</sup> See *Concord Music Grp., Inc. v. Anthropic PBC*, 23-cv-01092 (M.D. Tenn. Nov. 29, 2023).

<sup>84</sup> *Disney Enterprises, Inc. v. VidAngel, Inc.*, 224 F. Supp. 3d 957, 969 (C.D. Cal. 2016), *aff’d*, 869 F.3d 848 (9th Cir. 2017).

secure, access-controlled location in the cloud.”<sup>85</sup> Those facts appear to be quite similar to AI ingestion. The U.S. District Court for the Central District of California held the following:

VidAngel contends that case law regarding the reproduction right under § 106(1) draws a clear distinction between unlawful copies, which can be viewed by consumers, and lawful “intermediate” copies, which cannot be viewed. VidAngel argues that since their intermediate copies are unable to be viewed by consumers, they are not “copies” as defined by the Copyright Act and, as a matter of law, do not give rise to infringement claims. Defendants cite the Ninth Circuit’s decision in *Sega Enters. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992) as support for their proposition that “intermediate” copying does not violate the Copyright Act. However, the court in *Sega* stated that “on its face, the language of 17 U.S.C. § 106(1) unambiguously encompasses and proscribes ‘intermediate copying’” *Id.* at 1518. 17 U.S.C. § 101 provides that “in order to constitute a “copy” for purposes of the Copyright Act, the allegedly infringing work must be fixed in some tangible form, “from which the work can be *perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device.*” (emphasis added). VidAngel’s fragmented copies may not be able to be perceived directly by consumers, however they are able to be perceived with the aid of VidAngel’s software. Thus, the copying performed by Defendants falls within the category of acts that are proscribed by the statute.<sup>86</sup>

While commentators are correct to say that intermediate copying *can* be noninfringing,<sup>87</sup> it does not follow that an intermediate copy does not constitute a “copy,” as defined by the Copyright Act. Like in *VidAngel*, during the AI ingestion process, *expressive* elements are copied in a way that can be perceived by a machine or computer system. Further, similar to VidAngel, AI developers are making unauthorized copies, regardless of whether they are “intermediate” copies, are fragmented, or cannot be viewed by a user of the AI model. Thus, claims that

---

<sup>85</sup> *Id.*

<sup>86</sup> *Id.* at 969-70.

<sup>87</sup> Anthropic PBC, *supra* note 5, at 7.

intermediate copying does not result in a copy, or that generative AI involves intermediate copying that is never infringing, should be rejected.

## 8. Different AI Models Raise Unique Issues that Require Specific Analysis

When addressing questions of infringement and fair use related to the ingestion of copyrighted works for training purposes, most commentators tend to focus their analysis on Large Language Models (LLMs) that ingest written material and text.<sup>88</sup> Those commentators argue that LLM models only copy unprotectable data about the relationships and patterns amongst words as a means of generating new text based on a statistical estimation of a satisfactory response to a prompt and that massive amounts of material are needed to train LLMs.<sup>89</sup>

In our initial comments and above, we explain why these “relationships” drawn from literary works constitute copyrightable expression. To further emphasize this point and why copying of copyrighted works without a license on such a massive scale is unnecessary, consider generative AI technology alongside existing written word analytical AI technologies like spellcheck, grammar correction tools, and autocomplete. Generative AI developers claim that LLM models “evaluate[] the proximity, order, frequency, and other attributes of portions of words” and merely “analyze the structure and syntax of language.”<sup>90</sup> But analyzing attributes of words, structure, and syntax is something that existing spellcheck and autocomplete technologies have done for years. Those technologies typically work by analyzing and applying the underlying rules and patterns of language and compiling lists of incorrect language usage and writing for predictive or corrective purposes—without the need to copy *massive* amounts of copyrighted works. To the extent copyrighted works are used, based on our research, that is typically done with the

---

<sup>88</sup> See, e.g., OpenAI, *supra* note 55; Andreessen Horowitz (a16z), Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-9057>.

<sup>89</sup> See, e.g., Google, *supra* note 4, at 4–5, 9.

<sup>90</sup> *Id.* at 4; OpenAI, *supra* note 55, at 6–7.

permission of the owners of those works.<sup>91</sup> If LLM developers are only interested in analyzing things like structure and syntax, as commentators claim, there is no reason the LLM model cannot work in a similar manner to these earlier technologies by ingesting smaller corpora of works with authorization to do so. In other words, there would be no need for LLMs to copy massive amounts of copyrighted works to implement rules and patterns that these earlier technologies were able to implement without such massive and non-consensual copying of copyrighted works. The difference is that generative AI is manufacturing expressive output, and to accomplish that the model must copy and ingest expressive elements of works.

Another point worth mentioning is that it is misleading and unhelpful for commentators to only examine the legality of one specific generative AI model and then apply that analysis to other models. Depending on the type of generative AI, the model is designed, developed, and programmed differently, employs different technology, and ingests different types of material. For example, generative image tools employ diffusion technologies that LLMs do not, and the fair use and interim copying analyses can be vastly different because of the unique operations of different models and the distinctive qualities of the works they ingest.

Avoiding generalizations or across-the-board rules is just as important when considering similar categories of AI models (e.g., different types of LLMs) that may operate differently or are employed for distinct purposes. Consequently, the facts and analysis relevant to one type of generative AI model may be irrelevant or apply differently in the context of a different type of generative AI model. Again, as we explain in our initial comments, infringement and fair use analyses are highly fact-specific and each type of model raises unique issues that require independent analysis.

---

<sup>91</sup> For example, Grammarly’s terms of service grant the company a license to user content for many purposes, including “developing new products or features.” GRAMMARLY, <https://www.grammarly.com/terms> (last visited Nov. 30, 2023).



## 9. Making Works Publicly Available Does Not Indicate Permission to Scrape and Current Technical Measures to Opt-Out Are Ineffective

Many commentators address current technical measures that may be used to prevent scraping of copyrighted works that are publicly available on the internet and claim that they are an effective way for copyright owners to opt out of having their works scraped.<sup>92</sup> It is a grave misconception that the availability and accessibility of protected works on the internet equates to permission to take, scrape, copy, or otherwise exploit these works absent the copyright owner's express authorization.

Our initial comments discuss the problems with robots.txt so we will not reiterate them here. Instead, we make two points that we did not make in our initial comments. First, robots.txt is unworkable when works are licensed downstream and copyright owners are unable to control those downstream uses. A copyright owner could, in theory, require a licensee to employ robots.txt or similar protocols to prevent scraping of a licensed work, but it is difficult to enforce and often impractical. For example, where a licensee purchases a copy of an image and uses it in an advertisement, it would be wholly impractical for a licensor to mandate the licensee to employ a robots.txt-like protocol to prevent scraping of that image because the advertisement could not be effectively publicized, defeating the whole point of the use of the image in the first place.

Second, the use of robots.txt and opt outs is not a new issue. The proposed use of robots.txt as a solution was first considered and rejected in *Associated Press v. Meltwater* (discussed in more detail above). In response to Meltwater's argument "that AP impliedly granted Meltwater a license to use the Registered Articles when it did not require its licensees to employ on their websites robots.txt protocol to exclude web crawlers,"<sup>93</sup> the court found the following:

---

<sup>92</sup> See e.g., CCIA, *supra* note 42, at 11; Library Copyright Alliance, *supra* note 77, at 5; Stability AI Comment Letter on the U.S. Copyright Office's Notice of Inquiry on Artificial Intelligence and Copyright, at 15, 17 (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-8664>; Anthropic, *supra* note 5, at 5; OpenAI, *supra* note 55, at 10.

<sup>93</sup> *Associated Press v. Meltwater U.S. Holdings, Inc.*, 931 F. Supp. 2d 537, 563 (S.D.N.Y. 2013).

- “[T]he failure of AP’s licensees to employ the robots.txt protocol did not give Meltwater an implied license to copy and publish AP content. First, what Meltwater is suggesting would shift the burden to the copyright holder to prevent unauthorized use instead of placing the burden on the infringing party to show it had properly taken and used content.”<sup>94</sup>
- “[T]here is no fair inference, based simply on the absence of the robots.txt protocol, that there has been a meeting of the minds between the copyright owner and the owner of the web crawler about the extent of copying.”<sup>95</sup>
- “[There is] no evidence to suggest that robots.txt instructions are capable of communicating which types of use the copyright holder is permitting the web crawler to make of the content or the extent of the copying the copyright holder will allow.”<sup>96</sup>
- “There are also practical problems with Meltwater's argument in the event that AP and its licensees wanted to continue to permit search engines to visit their sites. AP is engaged in an ongoing licensing program that includes granting licenses that permit the scraping of AP content by web crawlers from online sources. Robots.txt protocol can be adopted to allow or disallow specific web crawlers. If Meltwater's argument were successful, with each change in the list of licensees AP and each of its licensees would have to update their robots.txt protocol to indicate which web crawlers had permission to visit each site's webpages.”<sup>97</sup>
- “It is fair to assume that most Internet users (and many owners of websites) would like crawlers employed by search engines to visit as many websites as possible, to include those websites in their search results, and thereby to direct viewers to a vast

---

<sup>94</sup> *Id.*

<sup>95</sup> *Id.*

<sup>96</sup> *Id.*

<sup>97</sup> *Id.* at 563-64.

array of sites. Adopting Meltwater's position would require websites concerned about improper copying to signal crawlers that they are not welcome.”<sup>98</sup>

- Neither *Field v Google*<sup>99</sup> or *Parker v Yahoo!*<sup>100</sup> “suggests that AP impliedly consented to the copying done by Meltwater because its licensees permitted search engines to crawl their sites. These two decisions principally discuss a website protocol that performs a different function than robots.txt.”<sup>101</sup>

## 10. AI Licensing is Possible

Comments from AI companies and supporters argue that licensing copyrighted works for AI ingestion would inevitably result in paltry payments to right holders and marginalize individual copyright owners, and thus they conclude that licensing is a very bad idea.<sup>102</sup> Importantly, the issue of whether a license is necessary and how to distribute the monies resulting from those licenses are two very different questions and the answer to the second question should not affect the answer to the first. Even where the royalties per work or per use are small, that does not mean that those services and other licensees should not have to pay any licensing fee. Moreover, with sufficient volume, even low “per use” royalty rates can add up to considerable money. The notion that licensing should not be required because royalties may be small would turn copyright, and many other licensing models, on its head.

---

<sup>98</sup> *Id.* at 564.

<sup>99</sup> *Field v. Google Inc.*, 412 F. Supp. 2d. 1106 (D. Nev. 2006).

<sup>100</sup> *Parker v. Yahoo!, Inc.*, No. CIV.A. 07-2757, 2008 WL 4410095 (E.D. Pa. Sept. 25, 2008).

<sup>101</sup> *Meltwater*, 931 F. Supp. 2d at 564.

<sup>102</sup> See e.g., a16z, *supra* note 88, at 10; Anthropic, *supra* note 5, at 10; Hugging Face, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright, at 11-12 (Oct. 30, 2023), available at <https://www.regulations.gov/comment/COLC-2023-0006-8969>; Samuelson, Sprigman, & Sag, *supra* note 41, at 27.

Furthermore, as the AI industry scales and is projected to add more than \$4 trillion to the global economy,<sup>103</sup> payments to copyright owners can proportionally accumulate and grow over time, just as we have seen with other digital technologies and services. These licensing payments would apply across the board from individual to enterprise copyright owners, bolstering competitiveness, not creating entrenched monopolies as some commentators fear. But AI companies need to come to the negotiating table instead of persisting that they need not license from anyone. Quite simply, an AI tool has no value without the copyrighted materials on which they are trained, and the AI tool operators should not profit at the expense of the copyright owners whose valuable content is an essential part of the value of the AI tool.

In this context, the U.S. Copyright Office should continue to support a copyright owner's fundamental right and ability to license, or not to license, their works, which is crucial to ensure that the incentives of the copyright law system are not degraded, and that copyrighted works are being properly valued. Many of the content industries currently derive most of their revenues from licensing, and these licensing revenues should not be categorized as ancillary revenues.

Lastly, it should be recognized that technologies have already been developed that allow AI developers accessing certain works to read embedded metadata and find the associated licensing terms surrounding AI-related uses like text and data mining. Specifically, the PLUS Coalition created a new metadata property dedicated to data mining that has been adopted by as a standard by the International Press Telecommunications Council (IPTC).<sup>104</sup> It allows rightsholders and distributors to offer up a license for data mining, and any system accessing a digital image file can read the embedded metadata and find the license offering. We raise this point to further demonstrate how copyright owners are working to ensure that licensing information is readily available to AI developers and that compliance is not impractical.

---

<sup>103</sup> James Manyika & Michael Spence, *The Coming AI Economic Revolution*, FOREIGN AFFAIRS (Oct. 24, 2023), <https://www.foreignaffairs.com/world/coming-ai-economic-revolution>.

<sup>104</sup> *Rights Holders Can Exclude Images from Generative AI with IPTC Photo Metadata Standard 2023.1*, INT'L PRESS TELECOMMS. COUNCIL, <https://www.iptc.org/news/exclude-images-from-generative-ai-iptc-photo-metadata-standard-2023-1/> (last visited Nov. 30, 2023).

## II. ADDITIONAL COMMENTS

### 1. AI Companies Are Not Immune from Secondary Liability Under “Sony Safe Harbor”

Some commentators invoke *Sony v. Universal* to suggest that AI companies should be immunized from secondary copyright infringement claims because their services are capable of substantial non-infringing uses.<sup>105</sup> One commentator compares generative AI services to “staple articles” like cameras or video recorders and argues that “[m]isuse of AI systems to infringe copyright, much like misuse of a VCR or computer to impermissibly replicate copyrighted content, is attributable to the user, not the manufacturer of the system being abused.”<sup>106</sup> But while any secondary liability claim will depend on the specific facts, there are clear distinctions between generative AI services and the video recording device at issue in *Sony*. Further, commentators ignore the fact that the Supreme Court in *MGM v. Grokster* (1) explained that *Sony* was “about liability resting on imputed intent,” (not “about liability on any theory”), and (2) made clear that “*Sony* did not displace other theories of secondary liability.”<sup>107</sup> Thus, simply because a product is capable of substantial lawful use does not absolve the producer of secondary liability, and arguments that generative AI companies should enjoy blanket immunity from secondary liability should be rejected.<sup>108</sup>

---

<sup>105</sup> See, e.g., Engine, Comment Letter on the U.S. Copyright Office’s Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), at 9, available at <https://www.regulations.gov/comment/COLC-2023-0006-9000> (“In instances where a user is directing an AI model to generate content that infringes existing copyrighted material and then uses that generated content in some way that’s not protected by fair use, the developer of the AI tool should not face contributory liability.”); Google, *supra* note 4, at 14 (“Excluding developers of generative AI systems from the Sony safe harbor would put all innovation in the field of machine learning at risk.”).

<sup>106</sup> CCIA, *supra* note 42, at 12.

<sup>107</sup> *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, at 915 (2005).

<sup>108</sup> See CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW (2023) (explaining that when a user is the direct infringer, “the AI company could potentially face liability under the doctrine of ‘vicarious infringement,’ which applies to defendants who have ‘the right and ability to supervise the infringing activity’ and ‘a direct financial interest in such activities’”), <https://crsreports.congress.gov/product/pdf/LSB/LSB10922>.

First, *Sony* involved the distribution of a device (not a service) that (1) had no trace of or connection to a copyrighted work at the point of sale and (2) the manufacturer had no right or ability to control after distribution. In the generative AI context, most AI models are offered to users after ingesting massive amounts of copyrighted works—a point that, separately, raises the possibility of direct infringement on the part of the AI developer. Whether a user-prompted output is infringing, that output in one form or another owes its origin to or is derived from the material the model ingested.<sup>109</sup> In a case where a user-prompted output is substantially similar or otherwise qualifies as a direct infringement of an ingested work, that infringement would not be possible *if not for the actions of the AI developer* that scraped and ingested the work without authorization. To that extent, generative AI tools are more comparable to a device or service that is “pre-loaded” with unauthorized copies of works, or that facilitates easy access to pirated works, and nothing like a recording device or camera that is distributed to a consumer with no connection to copyrighted works.<sup>110</sup>

Second, many AI companies control their service after it is offered to users and have a substantial continuing relationship with those users. AI companies may update their services, fine-tune them, add datasets, implement safeguards, provide support services, block access to a third-party, charge for access to their services, or any other number of actions that would qualify as an ability to control. This right and ability to control, which is a factor of vicarious liability, was not an issue in *Sony* or any other case where the relationship between the manufacturer and consumer ends at the point of sale. While the right and ability to control and supervise is not alone determinative of secondary liability, it is a characteristic of many online services, including generative AI, that makes *Sony* largely inapplicable.<sup>111</sup> In many cases involving peer-to-peer

---

<sup>109</sup> We used the term “derived” not in a derivative copyright law sense, but simply to recognize that the outputs of generative AI are necessarily the result of their inputs.

<sup>110</sup> *See Universal City Studios Prods. v. TickBox TV, LLC*, No. CV 17-7496, 2018 WL 1568698, at \*10 (C.D. Cal. Jan. 30, 2018) (finding that “by aggregating various unauthorized sources of copyrighted work and simplifying the process of accessing that work, [the preloaded device] undoubtedly enlarges that audience and thereby enlarges the scope of the infringement.”)

<sup>111</sup> *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1022 (9th Cir. 2001) (“The ability to block infringers' access to a particular environment for any reason whatsoever is evidence of the right and ability to supervise.”); *Capitol Records, LLC v. ReDigi Inc.*, 934 F. Supp. 2d 640, 660 (S.D.N.Y. 2013) (“Clearly, ReDigi Vicariously infringed Capitol's copyrights. As discussed, ReDigi exercised complete control over its website's content, user access, and sales.”).

file-sharing services that were found secondarily liable (or shut down because of liability concerns), the right and ability to control was a critical factor, regardless of actual or alleged capability of substantial noninfringing uses.<sup>112</sup> To be sure, there are factors that distinguish generative AI services from the file-sharing services of the past. However, generative AI shares more similarities with those services than with the distribution of a VCR or tape recorder, and there is no bright line rule that shields them from secondary liability.

To be clear, we are not suggesting that generative AI tools are not capable of substantial non-infringing uses, nor are we suggesting that an AI company would necessarily be secondarily liable for any direct infringements by its users. Those questions will be answered based on the facts of specific cases and may take into account steps AI developers take to prevent rampant infringement. We only urge the Copyright Office to recognize that any broad assertions about “*Sony* safe harbor” are misguided and that the devices at issue in *Sony* are clearly distinguishable from generative AI services.

## **2. Copyright Office Registration Guidance Should Avoid the Term De Minimis**

As we explain in our initial comments, there are inconsistencies between the recent guidance<sup>113</sup> and parts of the Copyright Office Compendium on registration guidelines that must be clarified. Specifically, the guidance says that AI-generated content that is more than de minimis should be explicitly disclaimed in an application. In comments made during the Office’s webinar on the guidance, it was explained that “more than de minimis” means that the portion of the work would be itself copyrightable if created by a human.

Use of the de minimis standard to separate material that should be disclaimed from material that should not be is not only confusing but also, we believe, an inappropriate standard. The term “more than de minimis” is used in other contexts in the law where it does not involve a standard

---

<sup>112</sup> *Id.* See also Alex Bracetti, *A History of P2P Sites Being Shut Down*, COMPLEX (Jan. 28, 2012), <https://www.complex.com/pop-culture/a/alex-bracetti/a-history-of-p2p-sites-being-shut-down>.

<sup>113</sup> Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16,190 (Mar. 16, 2023), <https://www.govinfo.gov/content/pkg/FR-2023-03-16/pdf/2023-05321.pdf>.

of separable copyrightability. Specifically, in the context of joint authorship, a more than de minimis contribution may suffice to constitute joint authorship while not rising to the level of independent copyright protection.<sup>114</sup> Further, whether a portion of a work used without authorization is “more than de minimis” may have different implications in an infringement analysis. Because the de minimis standard varies in different contexts, we recommend that the Office not use the term as the standard for determining when it is necessary to disclaim material in a registration application. Instead, we recommend that the Office’s guidance confirm the standard articulated in section 621.2 of the Compendium, which explains that “[u]nclaimable material should be disclaimed only if it represents an appreciable portion of the work as a whole.”<sup>115</sup>

### **3. The U.S. Should Not Adopt the Exceptions and Limitations Approach Taken by Other Countries**

Some commentators urge that the U.S. should consider enacting various AI-related exceptions and limitations in copyright law adopted by only a handful of countries, noting that our country could fall behind and be unduly harmed by both allies and adversaries due to these inconsistencies.<sup>116</sup> These risks are overstated and disregard the harm to our creative economy that would result from overbroad AI-related exceptions. As stated in our initial comments, we should not change U.S. copyright law to harmonize it with exceptions and limitations for AI that exist in other countries. The U.S. has long been the leader in the creation of new copyrighted works and new technologies. Our country is a cultural leader and spawned a creative economy that contributes more than \$2.9 trillion to the U.S. gross domestic product and workforce that employs nearly

---

<sup>114</sup> 1 NIMMER ON COPYRIGHT § 6.07[A][1] (2023) (citing *Picture Music, Inc. v. Bourne, Inc.*, 314 F. Supp. 640 (S.D.N.Y. 1970), *aff’d on other grounds*, 457 F.2d 1213 (2d Cir. 1972)).

<sup>115</sup> U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES § 621.2 (3d ed. 2021).

<sup>116</sup> *See, e.g.*, a16z, *supra* note 88, at 8 (“There is a very real risk that the overzealous enforcement of copyright when it comes to AI training—or the ad hoc limitation of the fair use doctrine that properly protects AI training—could cost the United States the battle for global AI dominance.”); OpenAI, *supra* note 55, at 13 (“A restrictive interpretation of fair use in the AI training context would put the U.S. at odds with this growing trend and could drive massive investments in AI research and supercomputing infrastructure overseas.”); *see also* Anthropic PBC, *supra* note 5, at 3.



16.1 million people.<sup>117</sup> This is not the time to relinquish that leadership role and begin following the approaches taken by other countries.

Even if following the lead of other countries was a good idea, which it is not, actually doing it is not an easy task. As discussed in more detail in our initial comments, the approaches taken by other countries are inconsistent with each other, oftentimes broadly and vaguely worded, likely to be contested in the courts, and risk noncompliance under the Berne Convention.

#### 4. AI and Humans “Learn” Differently

Many AI companies continue to anthropomorphize AI systems in their comments. In doing so, one of their goals is to make it seem like humans and AI “learn” in the same ways and that, since humans do not have to pay to learn, that AI should not either.<sup>118</sup> We explain in our initial comments the major differences between the way humans and AI “learn”—especially in the context of copyright law.<sup>119</sup> For example, unlike an AI system, humans do not produce exact copies of the copyrighted works, humans can easily forget things,<sup>120</sup> and most importantly humans are capable of original, organized, intellectual conceptions and thoughts and thus do not require learning from *all* copyrighted works in the world. Importantly, humans do not retain exact copies of works in large repositories that may be susceptible to hacking, which could result in extremely harmful infringement and piracy problems. Though rough analogies between AI and

---

<sup>117</sup> Robert Stoner & Jéssica Dutra, Copyright Industries in the U.S. Economy: The 2022 Report, INT’L INTELL. PROP. ALL. 8 (Dec. 2022), [https://www.iipa.org/files/uploads/2022/12/IIPA-Report-2022\\_Interactive\\_12-12-2022-1.pdf](https://www.iipa.org/files/uploads/2022/12/IIPA-Report-2022_Interactive_12-12-2022-1.pdf).

<sup>118</sup> See, e.g., Google, *supra* note 4; Anthropic PBC, *supra* note 5; Meta Platforms, Inc., *supra* note 4; Electronic Frontier Foundation, *supra* note 4; Hugging Face, *supra* note 102.

<sup>119</sup> NEWS MEDIA ALLIANCE, *supra* note 82 (explaining that GAI models “do not ‘learn’—they copy, and they do so on a massive scale that no human could replicate”).

<sup>120</sup> In our initial comments, we discussed the possibility of generative AI models “unlearning” material ingested for training. A point that we did not mention is that, as former Stability AI employee Ed Newton-Rex explains, “unlearning” or “untraining” a model is largely irrelevant because “[i]n a year’s time, no one will be using today’s models.” Instead, he explains that new models will be developed and “trained from scratch” with material that developers choose. The result being that “[i]f we agree not to train on copyrighted work, there’s no ‘unlearning’ to do.” Ed Newton-Rex (@ednewtonrex), X (formerly Twitter) (Dec. 2, 2023, 12:42 PM), <https://twitter.com/ednewtonrex/status/1731006008355647622>.

humans may be helpful to explain the mechanics of AI systems, when closely examining the application of infringement and fair use principles to AI mechanisms, these analogies are just that—rough approximations. They obscure the ways that AI models actually work and fail to justify the indiscriminate reaping of creative, expressive works for the commercial profit of AI companies at the expense of human creators, and crumble under closer examination under a proper application of copyright law principles.

One thing we did not address in our initial comments was this concept that humans do not pay to learn and therefore AI should not have to do so either. In fact, humans do pay to learn. A person who goes to a museum often pays an entrance fee. If that museum has no entrance fee, it's because the museum is supported by taxpayer dollars, and since we are all taxpayers, the museum-goer has indirectly paid the museum. Schools and libraries are likewise supported by taxpayer dollars. A college student pays for textbooks and tuition to get educated by professors. When we learn from content available on the internet we need to pay for internet, cellular, cable, or satellite to access those internet sites. And sometimes humans pay for the sites themselves either directly through subscription fees or through ads that appear on the page. We pay fees to see a documentary in the theater or to attend a play or musical performance. These are just a handful of examples. As much as we would like to think the acquisition of knowledge is costless, it simply is not. Further, as explained in comments by the News/Media Alliance, “no one is allowed to copy an underlying work just because they have an alleged good reason to read the underlying document but don't want to buy (or otherwise lawfully access) a copy.”<sup>121</sup> Ultimately, learning is not free for humans, and nor should ingestion (what some call “AI learning”) be free for generative AI systems.

---

<sup>121</sup> News Media Alliance, Comment Letter on the U.S. Copyright Office's Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), at 41, available at <https://www.regulations.gov/comment/COLC-2023-0006-8956> (“No human can rebut an infringement claim merely by showing that he has learned by consuming the works he copied, even if he puts this new knowledge to productive use later on . . . . A teacher who copies to broaden his personal understanding is a productive consumer, but he nonetheless must pay for the works he consumes. If the teacher's consumption of copyrighted works inspires him to create new scholarship, so much the better, but his subsequent productivity does not entitle him to a refund for the works that influenced him. In much the same way, machine learning makes consumptive use of copyrighted materials in order to facilitate future productivity. If future productivity is no defense for unauthorized human consumption, it should not excuse robotic consumption, either.” (quoting Benjamin L.W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 73-74 (2017))).

## 5. Fine Tuning

Finally, we would like to bring attention to the fact that much smaller datasets are being developed for fine tuning generative AI models through Low-Rank Adaptation (LoRA)<sup>122</sup> and Retrieval-Augmented Generation (RAG),<sup>123</sup> and they represent a growing threat to creators and copyright owners. In its initial comments to this study, the Graphic Artists Guild explains that AI image generator users have developed fine-tuning models by uploading as few as 20 works by an illustrator, and these models can in turn generate images that closely mimic a creator's style.<sup>124</sup> While some commentators argue that the inclusion of a particular work in a dataset has a negligible influence on a model, these custom datasets are much more likely to replicate protected expression. As described in our initial comments, not only is the use of these datasets more likely to produce outputs which replicate the original images, it has been reported that they have been used for retaliatory purposes against artists who have been vocal against the unauthorized use of their works for AI training.<sup>125</sup> As the creation and use of these type of datasets proliferates, it is essential that their propensity to infringe be recognized and that creators and copyright owners are able to enforce their rights.

---

<sup>122</sup> See *What is low-rank adaptation (LoRA)?*, Ben Dickson, TECHTALKS (May 22, 2023), <https://bdtechtalks.com/2023/05/22/what-is-lora/>.

<sup>123</sup> See *What Is Retrieval-Augmented Generation (RAG)?*, Alan Zeichick, ORACLE (Sept. 19, 2023), <https://www.oracle.com/artificial-intelligence/generative-ai/retrieval-augmented-generation-rag/>.

<sup>124</sup> Graphic Artists Guild, Comment Letter on the U.S. Copyright Office's Notice of Inquiry on Artificial Intelligence and Copyright (Oct. 30, 2023), at 3, available at <https://www.regulations.gov/comment/COLC-2023-0006-9054> (“[U]ser Laura\_De\_Martin generated images on a model fine-tuned on the artist Conrad Roset; his name is included in the set of text prompts provided. [] This model has been downloaded over 2,000 times from the Civitai platform, a platform that purports to ‘. . . offer an environment where users can upload, share, and discover custom models, each trained on distinct datasets. With the aid of AI media software, these models can be leveraged as innovative tools for crafting your own unique creations.’ <https://civitai.com/content/guides/what-is-civitai/>.”).

<sup>125</sup> *Id.* at 4 (citing to Dr. Ben Zhao's description of how Japanese artists were being retaliated against by having their images uploaded for fine-tuning) (Webinar, *Learn How to Protect Your Artistic Style from Mimicry*, GRAPHIC ARTISTS GUILD, at 53:00 (Sept. 20, 2022), <https://graphicartistsguild.org/product/protect-yourstyle-from-mimicry/>).

## **Conclusion**

We appreciate the opportunity to submit these reply comments and look forward to working with the Office and other stakeholders on these issues in the future.

Respectfully submitted,

A handwritten signature in black ink, appearing to read 'KK', with a long horizontal flourish extending to the right.

Keith Kupferschmid  
CEO  
Copyright Alliance  
1331 F Street, NW, Suite 950  
Washington, D.C. 20004

December 6, 2023