

Benjamin Gould (SBN 250630)
KELLER ROHRBACK L.L.P.
1201 Third Avenue, Suite 3400
Seattle, WA 98101-3268
Tel.: (206) 623-1900, Fax: (206) 623-3384
bgould@kellerrohrback.com

Counsel for Copyright Alliance

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

IN RE MOSAIC LLM LITIGATION

Master File Case No. 3:24-cv-01451-CRB

**BRIEF OF AMICUS CURIAE
COPYRIGHT ALLIANCE IN SUPPORT
OF PLAINTIFFS' MOTION FOR
PARTIAL SUMMARY JUDGMENT AND
PLAINTIFFS' OPPOSITION TO
DEFENDANTS' MOTION FOR
SUMMARY JUDGMENT**

Judge: Hon. Charles R. Breyer

TABLE OF CONTENTS

I. INTRODUCTION 1

 A. AI Training Does Not Serve a Transformative Purpose, but Even If It Did, a Finding of Transformativeness Does Not Control the Fair Use Analysis. 2

 B. Robust Licensing Markets for AI Training Exist and Market Harm Is Cognizable. 4

 C. The Fair Use Decisions Cited by Defendants Are Significantly Undermined by the Supreme Court’s Decision in *Warhol*. 6

 D. The Non-AI Cases on Which Defendants Rely Are Distinguishable. 7

 1. *Authors Guild v. Google, Inc.* 7

 2. Image Search Engine Cases 9

 3. *Google LLC v. Oracle America, Inc.* 10

 4. Reverse Engineering Cases 11

 E. Databrick’s Reliance on the *Kadrey* and *Bartz* Cases is Misplaced 12

 1. *Kadrey* 12

 2. *Bartz* 13

 F. Knowingly Sourcing Pirated Materials Should Weigh Heavily Against Fair Use 14

II. CONCLUSION 15

TABLE OF AUTHORITIES**Page(s)****Cases**

<i>American Geophysical Union v. Texaco, Inc.</i> , 60 F.3d 913 (2d Cir. 1994)	3
<i>Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith</i> , 598 U.S. 508 (2023)	1, 2, 3, 6, 7, 10, 13
<i>Authors Guild v. Google, Inc.</i> , 804 F.3d 202 (2d Cir. 2015)	7, 8, 9, 11
<i>Bartz v. Anthropic PBC</i> , 787 F. Supp. 3d 1007 (N.D. Cal. 2025).....	2, 12, 13, 15
<i>Campbell v. Acuff-Rose Music, Inc.</i> , 510 U.S. 569 (1994)	2, 3, 6, 7
<i>Fox News Network, LLC v. TVEyes, Inc.</i> , 883 F.3d 169 (2d Cir. 2018)	6
<i>Google LLC v. Oracle America, Inc.</i> , 593 U.S. 1 (2021).....	10, 11, 14
<i>Kadrey v. Meta Platforms, Inc.</i> , 788 F. Supp. 3d 1026 (N.D. Cal. 2025).....	2, 3, 4, 6, 12, 13
<i>Kelly v. Arriba Soft Corp.</i> , 336 F.3d 811 (9th Cir. 2003)	9, 10
<i>NXIVM Corp. v. Ross Inst.</i> , 364 F.3d 471 (2d Cir. 2004)	14
<i>Perfect 10, Inc. v. Amazon.com, Inc.</i> , 508 F.3d 1146 (9th Cir. 2007)	9
<i>Ringgold v. Black Ent. Television, Inc.</i> , 126 F.3d 70 (2d Cir. 1997)	5
<i>Sega Enterprises, Ltd. v. Accolade, Inc.</i> , 977 F.2d 1510 (9th Cir. 1992)	11, 12
<i>Sony Computer Entertainment, Inc. v. Connectix Corp.</i> , 203 F.3d 596 (9th Cir. 2000)	11
<i>VHT, Inc. v. Zillow Grp., Inc.</i> , 918 F.3d 723 (9th Cir. 2019)	7

Statutes

17 U.S.C. § 106(1)1

Other Authorities

AI Licensing by Copyright Owners, Copyright Alliance,
<https://copyrightalliance.org/artificial-intelligence-copyright/licensing/copyright-owners> (last visited Aug. 26, 2026).....4

Alex Reisner, *These 183,000 Books Are Fueling the Biggest Fight in Publishing and Tech*, The Atlantic (Sept. 25, 2023),
<https://www.theatlantic.com/technology/archive/2023/09/books3-database-generative-ai-training-copyright-infringement/675363/>.....14

Chris Stokel-Walker, *Cultural Barbarism: How AI Companies Are Destroying the World's Books*, The Telegraph (July 30, 2026),
<https://www.telegraph.co.uk/news/2026/07/30/why-ai-companies-are-destroying-millions-of-old-books/>5

Circuit to Warhol Estate: Google v. Oracle Does Not Dictate a Different Result, Pryor Cashman (Aug. 26, 2021), <https://www.pryorcashman.com/publications/circuit-to-warholestate-google-v-oracle-does-not-dictate-a-different-result> 10

Ed Nawotka, *Princeton University Press Partners with Cashmere on AI Licensing*, Publishers Wkly. (Aug. 3, 2026), <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/100971-princeton-university-press-partners-with-cashmere-on-ai-licensing.html>8

Getty Images and Perplexity Strike Multi-year Image Partnership, Getty Images (Oct. 31, 2025), <https://newsroom.gettyimages.com/en/getty-images/getty-images-and-perplexity-strike-multi-year-image-partnership>;.....5

Jiarui Liu, *An Empirical Study of Transformative Use in Copyright Law*, 22 Stan. Tech. L. Rev. 163, 185-86 (2019), available at <https://ssrn.com/abstract=3330236>6

Matt O'Brien, *ChatGPT-maker OpenAI Signs Deal with AP to License News Stories*, Associated Press (July 13, 2023), <https://www.ap.org/media-center/ap-in-the-news/2023/chatgpt-maker-openai-signs-deal-with-ap-to-license-news-stories/>8

Melville B. Nimmer & David Nimmer, *Nimmer on Copyright* § 13.05[A][4] (Matthew Bender, rev. ed. 2024)5

Michael M. Grynbaum & Cade Metz, *The Times and Amazon Announce an A.I. Licensing Deal*, N.Y. Times (May 29, 2025),
<https://www.nytimes.com/2025/05/29/business/media/new-york-times-amazon-ai-licensing.html>.....4

1	Neil Weinstock Netanel, <i>Making Sense of Fair Use</i> , 15 Lewis & Clark L. Rev. 715,	
2	754–55 (2011).....	6
3	Nina Corcoran, <i>Major Labels Sign Licensing Deals With AI Music Company Klay</i> ,	
4	Pitchfork (Nov. 20, 2025), https://pitchfork.com/news/warner-music-group-signs-	
5	licensing-deal-with-ai-music-company-klay	5
6	<i>Partnership with Axel Springer to Deepen Beneficial Use of AI in Journalism</i> , OpenAI	
7	(Dec. 13, 2023), https://openai.com/index/axel-springer-partnership	5
8	Shyamkrishna Balganesh, Jane C. Ginsburg & Peter S. Menell, <i>Comments on</i>	
9	<i>Preliminary Draft 9 of the American Law Institute's Restatement of Copyright</i> at 15	
10	(Sep. 26, 2023),	
11	https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=5177&context=fac-	
12	ulty_scholarship	6
13	<i>Supreme Court Affirms Andy Warhol's Prince Series Not Transformative Fair Use</i> ,	
14	Proskauer Rose LLP (June 14, 2023), https://www.proskauer.com/blog/supreme-	
15	court-affirms-andy-warhol-prince-series-not-transformative-fair-use	10
16	<i>Text and Data Mining License</i> , Elsevier, https://www.elsevier.com/about/policies-and-	
17	standards/text-and-data-mining/license (last visited Aug. 27, 2026).....	8
18	Tyler Ochoa, <i>U.S. Supreme Court Upholds Fair Use in Google-Oracle Software Battle</i>	
19	(<i>Guest Blog Post</i>), Technology & Marketing Law Blog (Apr. 8, 2021),	
20	https://blog.ericgoldman.org/archives/2021/04/u-s-supreme-court-upholds-fair-use-	
21	in-google-oracle-software-battle-guest-blog-post.htm	10
22	<i>Universal Music Group and Stability AI Announce Strategic Alliance to Co-Develop</i>	
23	<i>Professional AI Music Creation Tools</i> , Universal Music Group (Oct. 30, 2025),	
24	https://www.universalmusic.com/universal-music-group-and-stability-ai-announce-	
25	strategic-alliance-to-co-develop-professional-ai-music-creation-tools	5
26	<i>Universal Music Group to Transform Music Experience for Billions of Fans with</i>	
27	<i>NVIDIA AI</i> , Universal Music Group, (Jan. 6, 2026),	
28	https://www.universalmusic.com/universal-music-group-to-transform-music-	
	experience-for-billions-of-fans-with-nvidia-ai	4
	<i>Warner Music Group and Stability AI Join Forces to Build the Next Generation of</i>	
	<i>Responsible AI Tools for Music Creation</i> , Warner Music Group (Nov. 19, 2025),	
	https://www.wmg.com/news/warner-music-group-and-stability-ai-join-forces-to-	
	build-the-next-generation-of-responsible-ai-tools-for-music-creation	5

I. INTRODUCTION

The Copyright Alliance submits this amicus brief because our members have a strong interest in proper application of the Copyright Act, especially the application of copyright’s fair use doctrine to generative artificial intelligence (GAI). The Supreme Court has emphasized that fair use must be applied on a case-by-case basis and is dependent on the specific facts of a case. Prior fair use cases may be instructive, but they do not dictate the outcome of cases, even when they involve similar facts, uses, and/or technologies. Yet Defendants¹ seek to convince the Court otherwise and extract categorical holdings where there are none. Defendants argue that Mosaic’s copying² of copyright-protected material to train its GAI model “fits squarely within [a] well-established line of cases” that “recognize intermediate copying is fair use when it is a necessary step for a transformative use.” Defs.’ Mot. for Summ. J. (“Defs.’ MSJ”) at 14 (Dkt. 321). However, as other courts considering similar arguments in pending GAI infringement litigation have found, the cases upon which Defendants rely are distinguishable in a myriad of ways from the Defendants’ use, and thus should be afforded very limited weight. Moreover, in arguing that the allegedly transformative nature of its model should tip the scales in favor of fair use, Defendants improperly equate “transformative” in the literal sense with transformative use as delineated by the Supreme Court. In doing so, they disregard the Supreme Court decision in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023) [hereinafter *Warhol*], which clarified that even when a court finds a use to be transformative, that finding should have a limited effect on the ultimate fair use determination. Finally, Defendants’ reliance

¹ This brief will refer to Databricks and Mosaic as “Defendants” collectively and will name each entity separately where the distinction is relevant.

² Many AI companies frequently describe their activities as “scraping,” “ingesting,” “learning,” etc. This is mere euphemism, intended to obscure what they are actually doing: copying. Such activity violates the Copyright Act, absent the authorization of the copyright owner or a valid defense. See 17 U.S.C. § 106(1) (granting the copyright owner the exclusive right “to reproduce the copyrighted work in copies”).

1 on this District’s summary judgment orders in *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007 (N.D.
2 Cal. 2025) and *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026 (N.D. Cal. 2025), is misplaced,
3 as those cases demonstrate that the unauthorized use of copyrighted works for training will often not
4 qualify as fair use. This is especially true when (1) a robust licensing market exists for the use of
5 copyrighted works for AI training, and (2) AI developers knowingly source pirated material for
6 training, as Mosaic did. For these reasons, the Copyright Alliance respectfully submits that the Court
7 should deny Defendants’ motion for summary judgment.
8

9 **A. AI Training Does Not Serve a Transformative Purpose, but Even If It Did, a Finding of**
10 **Transformativeness Does Not Control the Fair Use Analysis.**

11 Central to the Supreme Court’s decision in *Warhol* was its confirmation that the defendant’s
12 purported “justification” must be considered as part of the first-factor analysis. This standard, which
13 was first explained by the Supreme Court’s discussion of parodic uses in *Campbell v. Acuff-Rose Music,*
14 *Inc.*, 510 U.S. 569 (1994), was unmistakably reaffirmed in *Warhol*, which confirmed that when an
15 original work and secondary use share the same or highly similar purposes, and the secondary use is
16 commercial, “a particularly compelling justification is needed.” *Warhol*, 598 U.S. at 547. Applying this
17 standard to the facts of this case, there is insufficient justification to support Defendants’ claims that
18 copying massive amounts of pirated works and ignoring existing licenses offered by copyright owners
19 was necessary to train its GAI models.
20

21 Defendants wrongly seek to disaggregate the output of GAI models from the training process
22 when describing the purpose of GAI. In describing the purpose of training, Defendants incorrectly claim
23 the copying “served a fundamentally different purpose from Plaintiffs’ books.” Defs.’ MSJ at 2.
24 Defendants’ motion ignores what comes after the initial training—most notably, the generation of
25 output that *serves the same purpose* of the ingested works. The *Warhol* decision and many other pre-
26 *Warhol* cases are careful to *consider the ultimate purpose of a use* and not simply end their analysis at
27 an intermediate step—whether that be reverse engineering, shrinking a work into a thumbnail image,
28

1 archiving a work for use in a plagiarism detection service or a searchable digital repository for those
2 with print disabilities, sampling, or training.

3 Demonstrating courts' considerations of the ultimate or actual purpose of a use, the Second
4 Circuit in *American Geophysical Union v. Texaco, Inc.*, 60 F.3d 913 (2d Cir. 1994), rejected over-
5 generalized fair use justifications for widespread infringement:
6

7 The purposes illustrated by the categories listed in section 107 refer primarily to the
8 work of authorship alleged to be a fair use, not to the activity in which the alleged
9 infringer is engaged. Texaco cannot gain fair use insulation for [its employee]'s archival
10 photocopying of articles (or books) simply because such copying is done by a company
11 doing research. It would be equally extravagant for a newspaper to contend that because
12 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.

13 *Id.* at 924.

14 *Warhol*, *Texaco*, and many of the cases discussed below make clear that, when considering the
15 purpose of the use in a factor-one fair use analysis, the analysis must not merely consider the
16 intermediate step but rather must take into account the ultimate purpose of the use. Thus, the purpose
17 of Defendants' MPT models cannot be considered in a vacuum of "training." "Training," standing
18 alone, is unlikely to constitute a sufficiently transformative use or purpose any more than "licensing"
19 standing alone was a transformative use or purpose in *Warhol* or "sampling" standing alone was a
20 transformative use or purpose in *Campbell*. Mosaic's copying of Plaintiffs' copyright-protected literary
21 works for training must be considered alongside its ultimate purpose: to generate output that serves the
22 same purpose and acts as a substitute for the ingested works.
23

24 Furthermore, even if the use is found to be for a transformative purpose, the Supreme Court
25 made clear in *Warhol* that whether a use is transformative is not dispositive of the question of fair use;
26 rather, it is merely one subfactor within the first statutory fair use factor. Indeed, as Judge Chhabria
27 recognized in *Kadrey*, "[t]here is certainly no rule that when your use of a protected work is
28

‘transformative,’ this automatically inoculates you from a claim of copyright infringement.” *Kadrey*, 788 F. Supp. 3d at 1035. Any factor-one analysis must involve a weighing of other considerations, such as the commercial nature of and justification for the use, both of which will factor prominently. Defendants’ MPT models are clearly part of a commercial venture, designed to bolster its position in the market, which weighs against fair use under the first factor and offsets in whole or in part any finding of transformative purpose. Therefore, even if Defendants’ use of copyrighted works for ingestion were found to be a transformative use in this case (though it is not), that would not necessarily mean factor one favors fair use.

B. Robust Licensing Markets for AI Training Exist and Market Harm Is Cognizable.

Defendants incorrectly argue that the fourth fair use factor weighs in its favor because the market for licensing Plaintiffs’ books for training is “hypothetical” and thus “not legally cognizable.” Defs.’ MSJ at 2. There is nothing speculative or theoretical about the robust and existing, ever-expanding licensing market for AI training. Since the public launch of GAI models a few years ago, there have been thousands of voluntary licenses directly entered into by copyright owners in connection with AI development.³ There are numerous examples of AI companies that are directly entering into licensing agreements with (and thereby compensating) copyright owners for the ingestion of their works, underscoring the viability and vibrancy of the voluntary AI licensing market supported by a robust copyright law framework.⁴

³ See *AI Licensing by Copyright Owners*, Copyright Alliance, <https://copyrightalliance.org/artificial-intelligence-copyright/licensing/copyright-owners> (last visited Aug. 26, 2026) (listing over 200 such licenses—a list that does not include numerous confidential agreements or those that are not publicly reported).

⁴ Recent examples of free-market licensing and partnership agreements include the ones between Universal Music Group and NVIDIA, *Universal Music Group to Transform Music Experience for Billions of Fans with NVIDIA AI*, Universal Music Group, (Jan. 6, 2026), <https://www.universalmusic.com/universal-music-group-to-transform-music-experience-for-billions-of-fans-with-nvidia-ai>; between the *New York Times* and Amazon, Michael M. Grynbaum & Cade Metz, *The Times and Amazon Announce an A.I. Licensing Deal*, N.Y. Times (May 29, 2025),

Defendants claim that because the plaintiffs themselves have not licensed their works for training, there is no harm under the fourth factor. Defs.’ MSJ at 20. But that argument ignores the law, which clearly does not require that a copyright owner has already entered a market at the time of infringement. The fourth factor considers harm to the actual or potential market, including traditional markets and those that are *reasonable or likely to be developed*.⁵ There is no question that, given the strong existing market for books for AI training, a market for plaintiffs’ works is reasonable and likely to be developed. Moreover, AI developers have made clear that they want to use every existing literary material for training, which would include plaintiffs’ books, and they are willing to go to great lengths to acquire everything they can.⁶

Defendants also argue that licensing markets “could not exist given the insurmountable transaction costs that any AI developer would face in attempting to individually license the massive

<https://www.nytimes.com/2025/05/29/business/media/new-york-times-amazon-ai-licensing.html>; between various news publishers and OpenAI, e.g., *Partnership with Axel Springer to Deepen Beneficial Use of AI in Journalism*, OpenAI (Dec. 13, 2023), <https://openai.com/index/axel-springer-partnership>; between Universal Music Group and Stability AI, *Universal Music Group and Stability AI Announce Strategic Alliance to Co-Develop Professional AI Music Creation Tools*, Universal Music Group (Oct. 30, 2025), <https://www.universalmusic.com/universal-music-group-and-stability-ai-announce-strategic-alliance-to-co-develop-professional-ai-music-creation-tools>; between Getty Images and Perplexity, *Getty Images and Perplexity Strike Multi-year Image Partnership*, Getty Images (Oct. 31, 2025), <https://newsroom.gettyimages.com/en/getty-images/getty-images-and-perplexity-strike-multi-year-image-partnership>; between Klay Vision, Inc. and Universal Music Group, Sony Music Group, and Warner Music Group, *Nina Corcoran, Major Labels Sign Licensing Deals With AI Music Company Klay*, Pitchfork (Nov. 20, 2025), <https://pitchfork.com/news/warner-music-group-signs-licensing-deal-with-ai-music-company-klay>; and between Warner Music Group and Stability AI, *Warner Music Group and Stability AI Join Forces to Build the Next Generation of Responsible AI Tools for Music Creation*, Warner Music Group (Nov. 19, 2025), <https://www.wmg.com/news/warner-music-group-and-stability-ai-join-forces-to-build-the-next-generation-of-responsible-ai-tools-for-music-creation>.

⁵ Melville B. Nimmer & David Nimmer, *Nimmer on Copyright* § 13.05[A][4], at 13-189 (Matthew Bender, rev. ed. 2024); see also *Ringgold v. Black Ent. Television, Inc.*, 126 F.3d 70, 81 (2d Cir. 1997).

⁶ Chris Stokel-Walker, *Cultural Barbarism: How AI Companies Are Destroying the World’s Books*, The Telegraph (July 30, 2026), <https://www.telegraph.co.uk/news/2026/07/30/why-ai-companies-are-destroying-millions-of-old-books/>.

amount of works necessary to train an LLM.” Defs.’ MSJ at 2. That is nothing more than the self-serving argument that was rejected by Judge Chhabria in *Kadrey*.

[T]he suggestion that adverse copyright rulings would stop this technology in its tracks is ridiculous. These products are expected to generate billions, even trillions, of dollars for the companies that are developing them. If using copyrighted works to train the models is as necessary as the companies say, they will figure out a way to compensate copyright holders for it.”

Kadrey, 788 F. Supp. 3d at 1036.

C. The Fair Use Decisions Cited by Defendants Are Significantly Undermined by the Supreme Court’s Decision in *Warhol*.

Many of the cases on which Defendants rely for its argument that its copying of Plaintiffs’ works is transformative were decided after *Campbell*, but before *Warhol*. During that three-decade period, courts took a very broad view of transformativeness, and in fact, a finding of transformativeness was very often dispositive of the ultimate fair use question.⁷ But *Warhol* dramatically changed the landscape, both narrowing what “transformative” means and clarifying that a finding of transformativeness does not lead ineluctably to a conclusion that the use was fair. Thus, all of the fair use cases on which Defendants rely, which were decided between *Campbell* and *Warhol*, and all of which relied heavily on a transformative use finding for their ultimate holding of fair use, simply do not reflect the current, post-*Warhol*, state of the law.⁸

⁷ Indeed, between 1994 and 2023, a determination that a use was transformative almost always meant that a court concluded it was fair use. A 2011 study found that of all the fair use cases decided in 2006 through 2010, the defendant won nearly 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 Jiarui Liu, *An Empirical Study of Transformative Use in Copyright Law*, 22 Stan. Tech. L. Rev. 163, 185–86 (2019), available at <https://ssrn.com/abstract=3330236> (finding that transformative use was the only statistically significant subfactor driving the first factor determination).

⁸ Shyamkrishna Balganesh, Jane C. Ginsburg & Peter S. Menell, *Comments on Preliminary Draft 9 of the American Law Institute’s Restatement of Copyright* at 15 (Sep. 26, 2023), https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=5177&context=faculty_scholarship

D. The Non-AI Cases on Which Defendants Rely Are Distinguishable.

Defendants wrongly argue that “[c]ourts have consistently held that it is fair use to make digital copies of protected works—including entire books—as an intermediate step to creating a transformative technology.” Defs.’ MSJ at 13. Importantly, the transformative nature and purpose of the uses in cases cited by Defendants are in no way analogous to the use and underlying GAI technology in the instant case.

1. *Authors Guild v. Google, Inc.*

While Defendants claim that *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015) [hereinafter *Google Books*] involves “analogous facts” to the instant dispute, it is distinguishable because it involved a completely different purpose for the use of the copyrighted materials—to provide information location services to drive readers to the relevant source material. Defs.’ MSJ at 11. Concluding that Google’s use was transformative, the appellate court held that “Google’s making of a digital copy to provide a search function . . . augments public knowledge by making available information *about* [p]laintiffs’ books without providing the public with a substantial substitute for matter protected by the [p]laintiffs’ copyright interests in the original works or derivatives of them.” *Google Books*, 804 F.3d at 207.

Significantly, the decision made clear that the case “tests the boundaries of fair use,” *id.* at 206—a position that the Ninth Circuit has agreed with—and may have come out differently had any one of several factors varied. *See VHT, Inc. v. Zillow Grp., Inc.*, 918 F.3d 723, 743 (9th Cir. 2019) (“We agree with the Second Circuit’s observation that the copyright dispute over the Google Books search engine ‘tests the boundaries of fair use.’” (citation omitted)). First, the Second Circuit explained that the fair

(“The Supreme Court’s *Warhol* decision shifted fair use analysis back to the course that Congress intended. It is important to recognize many cases decided during the interim period [between *Campbell* and *Warhol*] that boil fair use analysis down to a mere transformativeness inquiry are no longer good law.”).

1 use analysis would have been different if the purpose of Google’s scanning of literary works was to
 2 create substitutes for the original works. *See Google Books*, 804 F.3d at 214, 218–19. In other words,
 3 it was critical that Google used the books to provide information *about* the works and by serving as a
 4 “pointer” for readers to consume the original works. By contrast, Defendants cannibalize Plaintiffs’
 5 copyrighted works for the purpose of allowing its users to generate the same type of works and thus is
 6 much more likely to usurp the market for (and obviate any need to consume) Plaintiffs’ works.

8 Second, in *Google Books*, Google used the works solely *for the information* in the works, but
 9 in the instant case, Defendants are using the information and the *expressive elements* of the works.
 10 What Defendants consider to be unprotectable information is actually protected copyrightable
 11 expression.

12 Third, in *Google Books*, Google copied legitimate copies of books. But in the instant case,
 13 Mosaic copied copyrighted works from datasets known to contain *pirated* works.

14 Fourth, and perhaps most important, the court in *Google Books* concluded (with regard to the
 15 fourth fair use factor) that there was no actual or potential market for the licensing of copyrighted works
 16 to search engines. *Id.* at 226–27. This is significantly different than with GAI, where there is very
 17 clearly an actual market for the licensing of copyrighted works for ingestion.⁹
 18

19
 20
 21 ⁹ There is already high demand for corpora of copyrighted works for ingestion by AI systems, and
 22 copyright owners are offering and entering into various licensing agreements. Publishers and
 23 copyright owners of scientific and research works have either offered or entered into licensing
 24 agreements that allow for text and data mining (TDM) or other GAI-related uses. *See, e.g., Text and*
 25 *Data Mining License*, Elsevier, [https://www.elsevier.com/about/policies-and-standards/text-and-data-](https://www.elsevier.com/about/policies-and-standards/text-and-data-mining/license)
 26 [mining/license](https://www.elsevier.com/about/policies-and-standards/text-and-data-mining/license) (last visited Aug. 27, 2026); Ed Nawotka, *Princeton University Press Partners with*
 27 *Cashmere on AI Licensing*, Publishers Wkly. (Aug. 3,
 28 2026), [https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-](https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/100971-princeton-university-press-partners-with-cashmere-on-ai-licensing.html)
[news/article/100971-princeton-university-press-partners-with-cashmere-on-ai-licensing.html](https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/100971-princeton-university-press-partners-with-cashmere-on-ai-licensing.html).
 Multiple news organizations, including NewsCorp, the Associated Press, the *Atlantic*, and the
Financial Times, have reached deals with OpenAI for use of their works to train ChatGPT. *See, e.g.,*
 Matt O’Brien, *ChatGPT-maker OpenAI Signs Deal with AP to License News Stories*, Associated
 Press (July 13, 2023), [https://www.ap.org/media-center/ap-in-the-news/2023/chatgpt-maker-openai-](https://www.ap.org/media-center/ap-in-the-news/2023/chatgpt-maker-openai-signs-deal-with-ap-to-license-news-stories/)
[signs-deal-with-ap-to-license-news-stories/](https://www.ap.org/media-center/ap-in-the-news/2023/chatgpt-maker-openai-signs-deal-with-ap-to-license-news-stories/).

1 Any of these differences standing alone would likely be sufficient to push this case beyond “the
2 boundaries of fair use” set forth in *Google Books*. When taken together, there can be no doubt that they
3 do.

4 **2. Image Search Engine Cases**

5 Defendants cite the Ninth’s Circuit’s decisions in *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th
6 Cir. 2003) and *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007) as evidence that
7 courts “have consistently held that it is fair use to make digital copies of protected works—including
8 entire books—as an intermediate step to creating a transformative technology,” Defs.’ MSJ at 13, but
9 Defendants ignore the major distinction that in both cases there was no risk of supplanting the market
10 for the originals. In these cases, the court held that the defendants’ reproduction of the plaintiff’s photos
11 as thumbnail images qualified as fair use because the thumbnail images served an entirely different
12 purpose than the original images. In *Arriba*, the court held that the plaintiff’s photographs were “artistic
13 works intended to inform and to engage the viewer in an aesthetic experience,” in contrast to Arriba’s
14 use, which the court found was “unrelated to any aesthetic purpose” and that the search engine instead
15 “functions as a tool to help index and improve access to images on the internet and their related web
16 sites.” *Arriba*, 336 F.3d at 818.

17 Here, by contrast, both the purpose of Plaintiffs’ works that are ingested and the purpose of the
18 material generated by the MPT models are exactly the same: to “entertain, educate, provoke thought,
19 emotionally engage, and/or elicit empathy.” Defs.’ MSJ at 12. Because both the ingested work and AI-
20 generated output serve the same purpose, that “seriously weakens” the fair use claim. *Arriba*, 336 F.3d
21 at 819 (citation omitted).

22 The court also ruled in favor of fair use in *Arriba* because “[t]he thumbnails do not stifle artistic
23 creativity because they are not used for illustrative or artistic purposes and therefore *do not supplant*
24 *the need for the originals*.” *Id.* at 820 (emphasis added). This is very different from the generation of
25

literary material by Defendants’ AI model, where in many cases it is likely that such AI-generated output would “supplant the need for the original [i.e., the ingested work].” *Id.*

3. *Google LLC v. Oracle America, Inc.*

Defendants assert that the Supreme Court’s decision in *Google LLC v. Oracle America, Inc.*, 593 U.S. 1 (2021), also supports a finding of fair use, *see* Defs.’ MSJ at 13, but it is widely recognized that *Google*, which involved copying of functional computer code, has a very limited application—and is simply inapplicable when the work being used is a traditional copyrighted work (e.g., books, songs, movies, photographs, etc.).¹⁰ The decision in *Google* is expressly limited to the specific type of computer declaring code at issue in that case and cannot be applied to past or future fair use analyses outside of that specific context. *See Google*, 593 U.S. at 40 (clarifying that its decision “do[es] not overturn or modify [its] earlier cases involving fair use”).

¹⁰ *See, e.g.*, Brief for the United States as Amicus Curiae Supporting Respondents at 25, *Warhol*, 598 U.S. 508 (No. 21-869), 2022 WL 3574326 (“[Andy Warhol Foundation’s] reliance on *Google* is likewise misplaced. Emphasizing the difficulty of ‘apply[ing] traditional copyright concepts in th[e] technological world’ of ‘functional’ computer programs, the *Google* Court principally focused on the second statutory factor, emphasizing the copied code’s distance ‘from the core of copyright[.]’” (citation omitted)); *Circuit to Warhol Estate: Google v. Oracle Does Not Dictate a Different Result*, Pryor Cashman (Aug. 26, 2021), <https://www.pryorcashman.com/publications/circuit-to-warholestate-google-v-oracle-does-not-dictate-a-different-result> (“The Circuit explained how the Supreme Court ‘took pains to emphasize’ that the unique context of the [*Google*] case (software code) may make its holding less applicable in other contexts, especially in the context where the copyrighted material ‘serves an artistic rather than a utilitarian function.’”); *Supreme Court Affirms Andy Warhol’s Prince Series Not Transformative Fair Use*, Proskauer Rose LLP (June 14, 2023), <https://www.proskauer.com/blog/supreme-court-affirms-andy-warhol-prince-series-not-transformative-fair-use> (“Recently, the court in *Google v. Oracle* interpreted fair use broadly but limited its decision to the context of software codes.”); Tyler Ochoa, *U.S. Supreme Court Upholds Fair Use in Google-Oracle Software Battle* (Guest Blog Post), Technology & Marketing Law Blog (Apr. 8, 2021), <https://blog.ericgoldman.org/archives/2021/04/u-s-supreme-court-upholds-fair-use-in-google-oracle-software-battle-guest-blog-post.htm> (“Any Supreme Court opinion is important, and this one no doubt will be quoted often in future briefs and opinions. But other than clarifying the standard of review, I doubt the decision will have much impact on fair use cases that do not involve software.”).

1 Significantly, the Court was careful to distinguish computer code from highly expressive works
2 (like those at issue in this case) that have no functional elements that may impact a factor-two analysis,
3 stating that “computer programs differ from books, films, and many other ‘literary works’ in that
4 [software] programs almost always serve functional purposes.” *Id.* at 21. When considering the facts
5 and circumstances surrounding Defendants’ use of Plaintiffs’ creative, expressive works of authorship,
6 the instant case is easily distinguishable from the software-specific facts of *Google*.

7
8 Defendants also quote *Google* when arguing that fair use analyses must consider “significant
9 changes in technology.” Defs.’ MSJ at 10 (citation omitted). However, if technological progress and
10 innovation, standing alone, were the only litmus test for fair use, very few uses would not qualify.
11 While Defendants would no doubt prefer that result, the fair use test is much more nuanced and complex
12 than that. Mere innovation or creative progress is not the standard, nor should it be.

13 **4. Reverse Engineering Cases**

14
15 Defendants also wrongly rely on *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 203
16 F.3d 596 (9th Cir. 2000), in which the Ninth Circuit concluded that Connectix’s intermediate copying
17 for reverse engineering qualified as fair use, necessary to permit Connectix to make its non-infringing
18 game station function with PlayStation games. Specifically, the court held that Sony’s copyrighted
19 software code included functional elements that resulted in a lower degree of protection. Similar to the
20 decision in *Google*, the decision in *Sony* largely depended on the analysis of factor two and the work at
21 issue being a computer program. The factor-two analysis was central to the ultimate fair use finding,
22 and it would be inapplicable to the highly expressive works copied by Mosaic here.

23
24 The Ninth Circuit in *Sega Enterprises, Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992), a
25 case involving facts very similar to *Sony*’s, further confirmed the inapplicability of reverse engineering
26 computer code cases when the works at issue are highly expressive. The court explained that “[b]ecause
27 Sega’s video game programs contain unprotected aspects that cannot be examined without copying, we
28

1 afford them a lower degree of protection than more traditional literary works.” *Id.* at 1526. The works
2 that Mosaic copied are exactly the type of traditional literary works that the Ninth Circuit confirmed
3 deserve a higher degree of protection and thereby significantly alter any fair use analysis.

4 **E. Databricks’ Reliance on the *Kadrey* and *Bartz* Cases is Misplaced**

5 Very few GAI copyright infringement cases have been decided, but there are two from this
6 District that, while reaching different conclusions, demonstrate that in most cases the unauthorized use
7 of copyrighted works for training will not qualify as fair use.

8 **1. *Kadrey***

9 Defendants rely on *Kadrey* to argue that their use of copyrighted works is “highly
10 transformative” and should thus qualify as fair use. Defs.’ MSJ at 12 (citation omitted). However, as
11 discussed above, even if a use is found to be transformative under the first fair use factor, that is not
12 dispositive of the analysis.

13 Defendants cite *Kadrey* to argue that “[t]here is no serious question” that the purpose of the
14 MPT models is distinct from that of the ingested works because the purpose of the copyrighted works
15 “is to be read by humans in their entirety and thereby entertain, educate, provoke thought, emotionally
16 engage, and/or elicit empathy,” while the purpose of material generated by the model is “to generate
17 relevant, helpful responses to widely varied input prompts.” *Id.* (citation omitted). At a very high level
18 that may be one purpose of the output, but there is no doubt that another purpose is to generate output
19 that is used to entertain, educate, provoke thought, etc. For example, the stated purpose of the Mosaic’s
20 MPT StoryWriter model was “to read and write stories.” Compl. ¶ 39 (quoting Defendants) (Dkt.
21 No. 1). Generating stories would undoubtedly entertain, educate, and/or provoke thought in those who
22 read them, and thus serves the same purpose as copyrighted stories used to train the model.
23
24
25
26
27
28

1 **2. *Bartz***

2 Defendants cite the *Bartz* case to argue that its technology is “exceedingly” and “spectacularly”
3 transformative, which it claims favors fair use. Defs.’ MSJ at 10 (citation omitted). But in *Bartz*, Judge
4 Alsup erroneously applied a literal or common meaning of “transformative” instead of applying the
5 Supreme Court’s definition of “transformative use.” Under well-established case law and re-
6 emphasized in the Supreme Court’s recent decision in *Warhol*, a use is not transformative merely
7 because it produces something new and innovative. To be transformative, the Supreme Court confirmed
8 that a use must (1) have a further purpose or different character, (2) be “justified because copying is
9 reasonably necessary to achieve the user’s new purpose” and, (3) “further[] the goal of copyright,
10 namely, to promote the progress of science and the arts, without diminishing the incentive to create.”
11 *Warhol*, 598 U.S. at 531–32. In the *Bartz* case, Judge Alsup failed to apply this critical analysis, and to
12 take the same approach in this case would only serve to propagate the error.
13
14

15 Defendants also rely on a quote from the *Bartz* decision to analogize its actions to the non-
16 infringing actions of a human who reads and memorizes classic books and then “emulate[s] a blend of
17 their best writing.” Defs.’ MSJ at 13 (citation omitted). But that is not how computers work and is
18 merely an attempt to humanize the functions of a machine and avoid liability. Judge Chhabria explains
19 the critical distinction in the *Kadrey* decision:
20

21 [A]n LLM’s consumption of a book is different than a person’s. An LLM ingests text to
22 learn “statistical patterns” of how words are used together in different contexts. It does
23 so by taking a piece of text from its training data, removing a word from that text,
24 predicting what that word will be, and updating its general understanding of language
25 based on whether it was right or wrong—and then repeating this exercise billions or
26 trillions of times with different text. This is not how a human reads a book.

27 *Kadrey*, 788 F. Supp. 3d at 1044.

28 The reality is that GAI models ingest and reproduce perfect copies of copyrighted works (and
generate material) at a speed and scale that makes any comparison to human learning utterly inapt, and
Defendants’ attempt to liken the two should be soundly rejected.

F. Knowingly Sourcing Pirated Materials Should Weigh Heavily Against Fair Use

Defendants argue that works were not sourced for training from “shadow libraries,” and that the illicit origin of the Books3 dataset is irrelevant to an overall fair use analysis. Defs.’ MSJ at 14 (citation omitted). While the Books3 dataset may not be identical to shadow libraries like Anna’s Archive, Bibliotik, or other clearly illicit websites and services that distribute stolen works, it’s common knowledge that Books3 was built on pirated works derived from such entities.¹¹ AI developers, including Defendants, know or should have known that *Books3 contains massive amounts of pirated works*, and their choice to nonetheless use pirated works for training should weigh heavily against fair use.

While Defendants argue that “bad faith” is “legally irrelevant to the fair use inquiry,”¹² *id.* at 15, when AI companies knowingly use massive amounts of pirated copies of works without authorization, such use is not a mere act of bad faith; it’s an act that borders on criminal infringement that should weigh against fair use. The key difference between past cases that have considered bad faith and the current AI infringement cases is that those past cases involved the use of one or many *legitimate* copies, and that use was found to be fair use because the copy was not otherwise available. *See NXIVM Corp. v. Ross Inst.*, 364 F.3d 471, 478 (2d Cir. 2004) (finding fair use where defendants’ only access to the unpublished, non-public course manual came through an alleged breach of a non-disclosure agreement). In contrast, many of the pending AI training infringement cases involve the use of *millions of illegal copies*. Moreover, copyrighted works that AI companies wish to use for training are largely

¹¹ Alex Reisner, *These 183,000 Books Are Fueling the Biggest Fight in Publishing and Tech*, The Atlantic (Sept. 25, 2023), <https://www.theatlantic.com/technology/archive/2023/09/books3-database-generative-ai-training-copyright-infringement/675363/>.

¹² While Defendants cite *Google*, the Supreme Court made clear in that decision that it was not taking a position on whether good v. bad faith is a “helpful inquiry,” only that “given the strength of the other factors pointing toward fair use and the jury finding in Google’s favor on hotly contested evidence, that factbound consideration is not determinative in this context.” *Google*, 593 U.S. at 32-33.

1 available through legitimate sources, which is in stark contrast to past cases where works were not
2 available. The fact that AI companies’ intentional use of these illicit sites and services does not
3 presumptively disqualify them from claiming fair use is antithetical to the foundations and goals of our
4 copyright system and the rights it guarantees. Judge Alsup recognized this in *Bartz*, finding that
5 Anthropic’s sourcing of pirated works did not qualify as fair use and that “bad faith is not the basis for
6 this decision.” *Bartz*, 787 F. Supp. 3d at 1026.
7

8 **II. CONCLUSION**

9 The cases relied on by Defendants for their fair use arguments are of limited relevance here
10 given the narrowed interpretation of transformativeness post-*Warhol*, and the existence of a robust
11 licensing market for training and use of pirated works should weigh against fair use. This Court should
12 deny Defendants’ motion for summary judgment.
13

14 Respectfully Submitted,

15 DATED this 27th day of August, 2026.

16 KELLER ROHRBACK L.L.P.

17
18
19 By /s/ Benjamin Gould

Benjamin Gould (SBN 250630)
1201 Third Avenue, Suite 3400
Seattle, WA 98101-3268
Tel.: (206) 623-1900
Fax: (206) 623-3384
bgould@kellerrohrback.com

23 *Counsel for Amicus Curiae*
24 *Copyright Alliance*
25
26
27
28