Jacqueline C. Charlesworth (SBN 326742)
Ruby A. Strassman (SBN 346601)
FRANKFURT KURNIT KLEIN + SELZ PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
Telephone: (310) 579-9600
Facsimile: (310) 579-9650

Attorneys for Amicus Curiae
Association of American Publishers
(*Complete listing on signature page*)

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

(SAN FRANCISCO DIVISION)

RICHARD KADREY, et al.,

Individual and Representative Plaintiffs,

v.

META PLATFORMS, INC., a Delaware Corporation,

Defendant.

Case No. 23-cv-03417-VC

BRIEF OF *AMICUS CURIAE* THE ASSOCIATION OF AMERICAN PUBLISHERS IN SUPPORT OF PLAINTIFFS' MOTION FOR PARTIAL SUMMARY JUDGMENT AND OPPOSITION TO DEFENDANT'S MOTION FOR PARTIAL SUMMARY JUDGMENT

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

TABLE OF CONTENTS

Frankfurt Kurnit Klein+Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

TABLE OF AUTHORITIES

Page(s)

Cases

Statutes, Rules, & Legislative Materials

Other Authorities

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

STATEMENT OF INTEREST OF *AMICUS CURIAE*

Amicus curiae Association of American Publishers ("AAP") is the national trade association for book, journal, and education publishers in the United States. AAP's members include major commercial publishers of fiction and nonfiction; education publishers; small, specialized, and independent publishers; and nonprofit publishers such as university presses and scholarly research societies.

AAP's members have a direct and compelling interest in the efficacy, administration, and enforcement of federal copyright laws. They invest in and make available to the public a wide range of works, including literature, scholarship, professional content, and scientific journals. Just as the long-term public interest is served by protecting the exclusive rights of copyright owners, the long-term potential of AI technology will only be realized by preserving the marketable rights that enable authors, publishers, and AI developers to engage in mutually beneficial commercial transactions.[1]

SUMMARY OF ARGUMENT

Defendant Meta Platforms, Inc. ("Meta"), a company valued at over a trillion dollars, asks this Court to declare that it is free to appropriate and commercially exploit the content of copyrighted works on a massive scale without permission or payment for that content, a ruling that would have catastrophic consequences for authors and publishers of books, journals and other textual works protected by copyright. Meta asserts that the unauthorized appropriation of millions of human-authored works, including pirated libraries of books, to train its generative large language model ("LLM"), Llama, is a "quintessential" fair use. It is not.

Contrary to Meta's claims, there is no judicial precedent that condones the mass appropriation of copyrighted works to make use of their expressive content for commercial ends as a fair use of those works. Meta's claim of fair use is largely predicated on two false narratives. Seeking to establish that its exploitation is "transformative" under the first fair use factor of

[1] No party or their counsel authored this brief in whole or in part or contributed money intended to fund the preparation or submission of this brief.

section 107 of the Copyright Act,[2] Meta misleadingly portrays the LLM training process—in which works are systematically reproduced and ingested into Llama word by word—as merely recording "statistical information" about the works rather than capturing the content of the works.[3] Meta would have this Court believe that authors' original expression is not preserved in or exploited by the model. But this is not so. The LLM algorithmically maps and stores authors' original expression so it can be used to generate output—indeed, that is the very point of the training exercise.

Contrary to Meta's claims, there is nothing transformative about the systematic copying and encoding of textual works, word by word, into an LLM. It does not involve criticism or commentary, provision of a search or indexing utility, software interoperability, or any other purpose recognized as transformative under fair use precedents. Nor can Meta claim that the output of Llama offers commentary, searchability, or other functionality with respect to the works on which it is trained.

The second falsehood advanced by Meta is that developers like Meta have no choice but to rely on fair use because they are unable to obtain licenses for copyrighted materials to train their LLMs.[4] Seeking to avoid a finding of market harm under the critical fourth fair use factor,[5] Meta urges this Court to disregard numerous documented AI licensing agreements for textual works, going so far as to claim "no such market exists."[6] From this false premise Meta asserts it was entitled to help itself to millions of books and articles, conveniently stripped of digital rights management ("DRM") protections, from illicit sources.

Significantly, despite entering into discussions with book publishers to acquire authorized copies of their works to train Llama, Meta instead chose to acquire texts from notorious pirate

[2] 17 U.S.C. § 107(1) (court to consider the "purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes").
[3] *See, e.g.*, Def.'s Opening Br. ("Meta Br.") at 10, 22.
[4] *See, e.g.*, *id.* at 31-32.
[5] 17 U.S.C. § 107(4) (court to consider the "effect of the use upon the potential market for or value of the copyrighted work").
[6] Meta Br. at 3.

Frankfurt Kurnit Klein+Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

sites like LibGen and Anna's Archive.[7] In light of this history, it is perhaps unsurprising that Meta seeks to deny the very existence of a viable market for AI training materials. But Meta's claim is belied by numerous publicly announced deals entered into by copyright owners in response to the advent of generative AI in the public sphere to authorize the use of books and other texts by AI developers (as catalogued in the illustrative chart below).

This Court should reject Meta's assertion that its appropriation of copyrighted works to train Llama is transformative fair use. There is no legal precedent to support the view that the systematic encoding of copyrighted materials to exploit their expressive content is a transformative use under the first fair use factor. As the commercial purpose of Llama is not in dispute,[8] the first fair use factor strongly favors Plaintiffs.

This Court should also reject Meta's spurious assertion that there is no market for licensing of books and other textual works to AI developers. The existence of an active market for AI training materials is indisputable. A finding of fair use in this case could eviscerate the ability of authors and owners to receive compensation for the value of their copyrighted works and the exploitation of those works to build and operate LLMs, without which works Llama and other AI models would not and could not exist. The obvious harm to a robust and rapidly expanding market for AI training materials weighs decisively against fair use under factor four.

Finally, from a policy perspective, Meta's decision to appropriate millions of DRM-free books and other texts from pirate libraries is incompatible with a finding of fair use. A ruling that legitimizes such conduct in the name of fair use would be directly contrary to Congress' express intent when in enacting the Digital Millennium Copyright Act ("DMCA") in 1998, which updated the Copyright Act for the digital age.[9] In adopting the DMCA, Congress sought to ensure a thriving online marketplace for copyrighted works and those seeking to use them by safeguarding the ability of copyright owners to distribute their works in protected formats.[10] A determination of fair use in this case would directly undermine that objective by rewarding the intentional

[7] Pl.'s Opening Br. ("Plaintiffs' Br.") at 20-21; Meta Br. at 6-8.
[8] Meta Br. at 18.
[9] *See generally* DMCA, Pub. L. No. 105-304, 112 Stat. 2860 (1998).
[10] *See* S. Rep. No. 105-190, at 8 (1998).

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

exploitation of stolen works as an alternative to authorized access.

ARGUMENT

I. There Is No Fair Use Precedent That Condones the Mass Appropriation of Copyrighted Works to Exploit Their Expressive Content

As justification for its mass unauthorized exploitation of copyrighted works, Meta invokes earlier fair use precedents involving technologically driven copying, including *Authors Guild v. Google, Inc.* ("*Google Books*"),[11] *Authors Guild, Inc. v. HathiTrust* ("*HathiTrust*"),[12] *Kelly v. Arriba Soft Corporation*,[13] *Perfect 10, Inc. v. Amazon.com, Inc.*,[14] *A.V. v. iParadigms, LLC* ("*iParadigms*"),[15] *Sega Enterprises Ltd. v. Accolade, Inc.*,[16] *Sony Computer Entertainment, Inc. v. Connectix Corp.* ("*Sony Computer*"),[17] and *Google, LLC v. Oracle America, Inc.* ("*Oracle*").[18] None of these decisions, however, sanctions the appropriation and exploitation of copyrighted works to capitalize on their expressive content, as Meta is doing here.

There was no general declaration in either *Google Books* or *HathiTrust* that mass reproduction of copyrighted works to construct a product predicated upon large-scale copying has any presumptive claim to fair use. To the contrary, the *Google Books* panel was careful to cabin its holding to the particular circumstances before it, including the fact that Google's search functionality returned only snippets of text that did not permit meaningful consumption of expressive content.[19] Although Google made full-text copies of the books, it was not seeking to capitalize on, or allow users to exploit, the intrinsic value of those works. Even so, the court considered Google's copying to "test the boundaries of fair use," a sentiment that the Ninth Circuit shares.[20] Indeed, the court pointedly observed that had Google permitted users greater access to

[11] 804 F.3d 202 (2d Cir. 2015).
[12] 755 F.3d 87 (2d Cir. 2014).
[13] 336 F.3d 811 (9th Cir. 2003).
[14] 508 F.3d 1146 (9th Cir. 2006).
[15] 562 F.3d 630 (4th Cir. 2009).
[16] 977 F.2d 1510 (9th Cir. 1992)
[17] 203 F.3d 596 (9th Cir. 2000).
[18] 593 U.S. 1 (2021).
[19] *See Google Books*, 804 F.3d at 224-25.
[20] *Id.* at 206; *see also VHT, Inc. v. Zillow Grp., Inc.*, 918 F.3d 723, 743 (9th Cir. 2019) ("We agree

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

"the expressive content" of the book, such exploitation "would most likely constitute copyright infringement if not licensed by the rights holders."[21]

The determination of fair use in *HathiTrust* was similarly confined to the facts of that case, which also involved large-scale unauthorized scanning of books to create a searchable database.[22] In both *Google Books* and *HathiTrust*, the court distinguished between uses that were merely functional in nature and uses that sought to capitalize on expressive authorship. Meta's conduct cannot be squared with the fair use finding in either of these cases, the holdings of which were careful to preserve copyright owners' legitimate interest in their expressive authorship.

In *Kelly*, the court held that the defendant's copying of photos to provide a search and indexing service was a transformative fair use because the low-quality thumbnail copies "serve[d] a different function" than the originals by "improving access to information on the internet."[23] The court pointedly distinguished this purpose from copying to capitalize on "artistic expression."[24] Like *Kelly*, *Perfect 10* involved indexing of online images.[25] The Ninth Circuit again held that a search engine's copying of images for thumbnail display was a transformative fair use—and again because the images were not being used for their intrinsic purpose, but rather to create "an electronic reference tool."[26] Similarly, in *iParadigms*, which considered a plagiarism detection service that made copies of student papers, the Fourth Circuit focused on the fact that the defendant's use of the copied content "had an entirely different function and purpose than the original works," emphasizing that the use was "unrelated to any creative component" of the student works.[27]

Sega, *Sony Computer*, and *Oracle* are even farther afield from the type of copying engaged in by AI companies. In each of these cases, the copying was of a particular work to access non-

---

with the Second Circuit's observation that the copyright dispute over the Google Books search engine 'tests the boundaries of fair use.'") (quoting *Google Books*, 804 F.3d at 206).

[21] *Google Books*, 804 F.3d at 226.

[22] *HathiTrust*, 755 F.3d at 97-104.

[23] *Kelly*, 336 F.3d at 818-19.

[24] *Id.* at 819.

[25] *Perfect 10*, 508 F.3d at 1157.

[26] *Id.* at 1164-65.

[27] *iParadigms*, 562 F.3d at 639, 641-42.

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

protected computer code for the purpose of facilitating interoperability (in *Oracle*, interoperability among software coders)—clearly not the objective of Meta's copying. Notably with respect to the case before this Court, in *Sega*, although the Ninth Circuit concluded that reverse engineering of Sega's computer code to determine its functional elements to create compatible videogames was a fair use, the court expressly *rejected* the defendant's argument that its copying of Sega's code could not infringe because its end product was not substantially similar to the copied work. The *Sega* court ruled that copying "may infringe the exclusive rights granted to the copyright owner in Section 106 of the Copyright Act *regardless of whether the end product of the copying infringes those rights*."[28] This principle was reaffirmed by the Ninth Circuit two decades later in *Sony Computer*.[29]

In short, none of the fair use precedents on which Meta purports to rely addressed a product designed to copy and exploit authors' expressive content to derive new content, including potentially infringing or competing content. Not one involved the appropriation of expressive content for its intrinsic value. There is no judicial precedent that supports Meta's claim of fair use in this case.

II. Meta's Mass Appropriation and Exploitation of Expressive Works to Train Llama Was Not a Fair Use

Meta's mass copying of books and other textual works from illicit sources to build and operate Llama was commercially motivated and not a transformative use of those works. Accordingly, the first fair use factor weighs against fair use. With respect to the critical fourth fair use factor, market harm, Meta's choice to rely on stolen content rather than acquire licenses from publishers undermined a vital market for copyright owners seeking to license their works for AI training purposes. Factor four, then, also weighs heavily against fair use. As it is undisputed that Plaintiffs' works are creative (fair use factor two) and were taken in their entirety (factor

[28] *Sega*, 977 F.2d at 1517-19 (emphasis added). The court instead grounded its fair use finding on the fact that the copying was undertaken solely to identify functional elements rather than to exploit Sega's creative expression. *Id.* at 1522-23.

[29] *See Sony Computer*, 203 F.3d at 602-03 ("In *Sega*, we recognized that intermediate copying could constitute infringement even when the end product did not itself contain copyrighted material.").

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

three), the four factors together point decisively against fair use.

A. LLM Training Is Not a Transformative Use of Copyrighted Works

1. Training Is Not "Learning" About Works, But the Encoding of Expressive Content Word by Word

Meta claims that the LLM training process, by which copyrighted works are systematically ingested and encoded into the model word by word, is transformative because the LLM learns only information about the works, with "'nothing of the training data remain[ing].'"[30] According to Meta, Plaintiffs' works were appropriated and used merely to "extract … unprotected statistical data regarding word order, frequencies, grammar, and syntax," rather than "protected expression."[31] This characterization of the training process is grossly misleading.

To begin with, common sense dictates that authors' words themselves, not just "statistical information" about them, are stored in the model. Otherwise how could the model capture "word order" or "syntax"? And how would Llama generate word-based output?

The reality is that, to train an LLM, authors' expression is first copied wholesale from an online or other source—in Meta's case, by downloading texts from pirate sites—and then mapped word by word into the model so the model can draw upon that expression (and other authors' expression) to generate output. Textual works are broken down into small segments, or "tokens," typically consisting of a word or part of a word.[32] The tokens are encoded into word vectors, long number sequences that capture where the tokens appear in relation to other tokens in the text, so the text is represented in numerical form.[33] The vectorized tokens can be decoded and translated

---

[30] Meta Br. at 10 (quoting Meta expert Ungar).

[31] *Id*. at 22.

[32] *See* Meta Br. at 5; *see also* Lark Editorial Team, *Tokens in Foundational Models*, Lark (Dec. 25, 2023), https://www.larksuite.com/en_us/topics/ai-glossary/tokens-in-foundational-models; Hakan Tekgul, *Tokenization: Unleashing the Power of Words*, Arize (Feb. 2, 2023), https://arize.com/blog-course/tokenization/; Amal Menzli, *Tokenization in NLP: Types, Challenges, Examples, Tools*, neptune.ai Neptune Blog (Aug. 11, 2023), https://neptune.ai/blog/tokenization-in-nlp ("Menzli").

[33] *See* Menzli ("The token occurrences in a document can be used directly as a vector representing that document."); Babis Marmanis, *Heart of the Matter: Demystifying Copyright in the Training of AIs*, Dataversity (Feb. 2, 2024), https://www.dataversity.net/heart-of-the-matter-demystifying-copying-in-the-training-of-llms/("Marmanis"); AWS, *What are Large Language Models (LLM)?*,

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

into text again.[34] Within the model, these vectorized representations of the work's content, also known as "embeddings," are used for the model's generative activities.[35] As explained by a software engineer:

> The [] vectors are representations of tokens that preserve their original natural language representation that was given as text. It is important to understand the role of word embeddings when it comes to copyright because the embeddings form representations (or encodings) of entire sentences, or even paragraphs, and therefore, in vector combinations, even entire documents in a high-dimensional vector space. It is through these embeddings that the AI system captures and stores the meaning and the relationships of words from the natural language.[36]

In other words, authors' copyright-protected expression is encoded and stored in the model by Meta and other AI developers.

2. Meta's Use of Copyrighted Works to Train Llama Was Not Transformative

The process described above is not a transformative use of copyrighted works.

Looking to the initial stage of the training process, Meta does not dispute that it engaged in unlicensed full-text copying when, using torrenting technology, it downloaded millions of textual works from illicit sources to assemble training sets for Llama.[37] That Meta's wholesale appropriation of texts constituted direct copying without any type of transformation seems obvious.

Nor was there anything transformative about Meta's mechanical encoding of the works it appropriated word by word into Llama. The mechanical mapping of the content of copyrighted works to exploit that content for its intrinsic expressive value (as opposed merely to enabling a searching or indexing function, for example) does not qualify as a transformative use. It is well established that encoding a copyrighted work into a more convenient or usable format is an act of

---

https://aws.amazon.com/what-is/large-language-model/ (last visited Apr. 11, 2025) ("AWS"); Kevin Henner, *An intuitive introduction to word embeddings*, Stack Overflow Blog (Nov. 9, 2023), https://stackoverflow.blog/2023/11/09/an-intuitive-introduction-to-text-embeddings/.

[34] Janakiram MSV, *The Building Blocks of LLMs: Vectors, Tokens and Embeddings*, The New Stack (Feb. 8, 2024), https://thenewstack.io/the-building-blocks-of-llms-vectors-tokens-and-embeddings/.

[35] *Id.*; AWS.

[36] Marmanis.

[37] Meta Br. at 6-8, 13.

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

copying that does not itself qualify as transformative under the criteria for fair use.[38] In the words of the Ninth Circuit, courts are "reluctant to find fair use" when a work is merely converted to another format or medium.[39]

Rather than focusing on the mechanical word-by-word copying of works to populate its LLM, Meta argues that the "astonishing" and "cutting-edge" capabilities of Llama[40] to generate new content render that copying transformative.[41] This is the equivalent of arguing that the wondrous ability of an online music service to encode and store tens of millions of songs so they can be called up and streamed back at a user's command renders the service's exploitation of music transformative (which of course it is not).

But even if one considers Meta's claim that Llama's ability to generate output renders its mass copying transformative, its claim to transformativeness fails. The use of copyrighted works to facilitate AI generation does not align with the reasoning of *Google Books*, *HathiTrust*, or other technological cases in which the copying was found to be transformative. As explained above, in each of the cases relied upon by Meta the copying was to facilitate a use of the copyrighted work or works that did not aim to capitalize on expressive authorship—such as to provide a search or indexing service, or enable interoperability. By contrast, the purpose of Meta's copying was to exploit the expressive content of the works, thus supplanting an obvious licensing market for those works (as discussed below).

As Meta acknowledges, Llama can and does generate infringing copies of the works on

[38] *A&M Records, Inc. v. Napster*, 239 F.3d 1004, 1015 (9th Cir. 2001) ("*Napster*"); *see also, e.g.*, *Hachette Book Grp. v. Internet Archive*, 115 F.4th 163, 192 (2d Cir. 2024) (digitizing books not transformative for purposes of fair use); *Disney Enters, Inc. v. VidAngel, Inc.*, 869 F.3d 848, 861-63 (9th Cir. 2017) (rejecting argument that encoding of motion pictures to operate a streaming service was a transformative fair use); *UMG Recordings, Inc. v. MP3.com, Inc.*, 92 F. Supp. 2d 349, 351 (S.D.N.Y. 2000) ("*MP3.com*") (same for music); U.S. Copyright Office, *Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies*, 80 Fed. Reg. 65944, 65960 (Oct. 28, 2015) (rejecting notion that format-shifting or space-shifting constitutes a fair use).

[39] *Napster*, 239 F.3d at 1015.

[40] Meta Br. at 1, 4.

[41] *Id.* at 15, 17, 19-20.

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

which it is trained.[42] To state the obvious, an AI model's generation of text or an image that reproduces a training work is not transformative.[43] Nor is there anything inherently transformative about combining elements of one work with those of another work or works, which invades the copyright owners' derivative work rights.[44] As Professor Jane Ginsburg observes, "AI outputs may incorporate the source works' expression in a new production; but that output generally will not comment, criticize, shed light on or otherwise be *about* the copied expression."[45]

In *Warhol*, the Supreme Court warned against "an overbroad concept of transformative use" that encroaches upon copyright owners' derivative work rights, explaining that an interpretation of transformativeness "that includes any further purpose, or any different character" could "swallow" the copyright owner's exclusive right to create derivative works.[46] To this end, the Court criticized overzealous application of "transformativeness" to encompass any work that "adds some new expression, meaning, or message."[47] Drawing on its earlier explication of fair use in *Campbell v. Acuff-Rose Music, Inc.*,[48] the Court emphasized that the secondary user must have an independent justification for use of the work in question; that a copied work may be useful to convey a new meaning or message is not justification enough.[49]

---

[42] Meta Br. at 9, 23-24.

[43] *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 532-33 (2023) ("*Warhol*") (first fair use factor likely to weigh against fair use where "an original work and a secondary use share the same or highly similar purposes, and the secondary use is of a commercial nature").

[44] *See, e.g., id.* at 537, 550-51 (unlicensed commercial use of plaintiff's photograph, as incorporated into an Andy Warhol silkscreen derivative, was nontransformative and therefore infringing); *Dr. Seuss Enters., L.P. v. ComicMix LLC*, 983 F.3d 443, 451-55 (9th Cir. 2020) ("*Seuss*") (ComicMix's unlicensed book consisting of a "mashup" of Dr. Seuss and Star Trek characters that mimicked Dr. Seuss illustrations was a nontransformative use of Seuss's works).

[45] Jane C. Ginsburg, *Fair Use in the US Redux*, Singapore J. Legal Stud. 3 at 29 (Mar. 2024) (emphasis in original), https://law.nus.edu.sg/sjls/wp-content/uploads/sites/14/2024/05/firstview-march24-JaneGinsburg.pdf.

[46] *Warhol*, 598 U.S. at 529, 541. In keeping with this instruction, the Court determined that a magazine's commercial use of a silkscreen image created by Andy Warhol from plaintiff Goldsmith's photographic portrait of Prince was not transformative because it served as a substitute for Goldsmith's original photo. *Id.* at 523-24.

[47] *Id.* at 541.

[48] *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994).

[49] *See Warhol*, 598 U.S. at 532-33 (independent justification "particularly relevant" where

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

AI-generated content that is not a recognizable copy or derivative of a training work or works—that is, the type of content AI companies claim to be the intended output of their systems[50]—by definition does not comment or shed light on any particular work. It is difficult to see how Meta can stake a claim to transformative use of training works based on output that does not convey commentary or criticism with respect to the particular works it has copied.

B. Meta's Use of Stolen Works Undermines a Critical and Rapidly Expanding Licensing Market for the Use of Textual Works by AI Developers

There is one key point on which Plaintiffs agree with Meta: long-form textual works such as books are extremely valuable as training material for LLMs.[51] Books and journals are rich in high-quality, structured, long-form prose and undergo a rigorous editorial process, ensuring grammatical and factual accuracy.[52] Moreover, they span a range of topics, disciplines, and genres, and can thereby supply LLMs a wide variety of linguistic styles and concepts.

Claiming "there is no evidence that a market for licensing books to train LLMs" exists, and "none of Plaintiffs' work has economic value," Meta asserts that its unlicensed copying to train Llama has not harmed any cognizable market for Plaintiffs' works.[53] Meta further proclaims that "any theoretical market for licensing text as training data is doomed to 'market failure' as there is "no economically feasible mechanism for Meta or other LLM developers to obtain licensed copies."[54] Meta's claims are patently false. Since AI emerged in public life with the launch of ChatGPT at the end of 2022, AI companies including OpenAI (the company behind ChatGPT), Microsoft, Amazon, and others have entered into content licensing deals with publishers in order

---

unlicensed copying could displace a market for derivatives), 532-33 ("If an original work and a secondary use share the same or highly similar purposes, and the secondary use is of a commercial nature, the first factor is likely to weigh against fair use, absent some other justification for copying."), 547 ("Copying might [be] helpful to convey a new meaning or message. It often is. But that does not suffice under first factor.").

[50] *See, e.g.*, Meta Br. at 23-34 (acknowledging Meta has had to "implemen[t] mitigations" to "guard against the possibility of infringing outputs").

[51] Plaintiffs' Br. at 4-6; Meta Br. at 23.

[52] Plaintiffs' Br. at 4-6.

[53] *See* Meta Br. at 26-32.

[54] *Id.* at 30-31.

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

to access and use their works to build and operate AI systems. The below chart lists AI licensing deals for textual works of which AAP is aware, either directly from its members or through public reports. Undoubtedly there are many more that are not known to AAP or are still in the pipeline.

AI Licensing Deals for Textual Works
(As Publicly Announced and/or Reported by AAP Members)[55]

Licensee	Content Owner
Amazon	The Associated Press
Amazon	Business Insider
Amazon	Condé Nast
Amazon	Forbes
Amazon	Hearst
Amazon	Politico
Amazon	Reuters
Amazon	Time
Amazon	USA Today
Amazon	The Washington Post
Amazon	Vox
Dow Jones	The Associated Press
Dow Jones	The Wall Street Journal
Dow Jones	The Washington Post
LexisNexis	The Associated Press
Meta	Reuters
Microsoft	Axel Springer
Microsoft	Financial Times
Microsoft	HarperCollins
Microsoft	Hearst
Microsoft	Reuters
Microsoft	USA Today
Mistrial	Agence-France-Press
OpenAI	American Journalism Project
OpenAI	The Associated Press
OpenAI	The Atlantic
OpenAI	Axel Springer
OpenAI	Axios
OpenAI	Condé Nast
OpenAI	Dotdash Meredith
OpenAI	Financial Times
OpenAI	GEDI
OpenAI	Guardian Media Group

Licensee	Content Owner
OpenAI	Hearst
OpenAI	Le Monde
OpenAI	News Corp
OpenAI	Prisa Media
OpenAI	Schibsted Media Group
OpenAI	TIME
OpenAI	Vox Media
Perplexity	Adweek
Perplexity	The Independent
Perplexity	The Los Angeles Times
Potato	Wiley
ProRata.ai	Adweek
ProRata.ai	The Atlantic
ProRata.ai	Atlas Obscura
ProRata.ai	Arena Group
ProRata.ai	Axel Springer
ProRata.ai	Buzzfeed
ProRata.ai	DMG Media Group
ProRata.ai	Financial Times
ProRata.ai	Fortune
ProRata.ai	Guardian Media Group
ProRata.ai	Hello!
ProRata.ai	Mediahuis
ProRata.ai	Mumsnet
ProRata.ai	News/Media Alliance (on behalf of its members)
ProRata.ai	Prospect
ProRata.ai	Reach PLC
ProRata.ai	Sky Media Group
Confidential	Taylor & Francis
Confidential	Wiley
Confidential	Wiley
Confidential	Wiley
Confidential	Wiley

[55] Information used to compile this chart is on file with AAP.

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

The fourth fair use factor of section 107 requires courts to assess the effects of the claimed fair use on the "potential market" for the works at issue. This includes "not only the extent of market harm caused by the particular actions of the alleged infringer, but also *whether unrestricted and widespread conduct of the sort engaged in by the defendant ... would result in a substantially adverse impact*."[56] As illustrated above, the market for licensing of textual works to AI developers is not hypothetical, but actual—and rapidly expanding. Some researchers estimate the AI training license market to be valued at $2.5 billion now, *projecting it to reach $30 billion within a decade*.[57]

Licensing structures continue to evolve that enable authors and content owners to participate in collective deals, receive attribution when AI tools rely upon their work, and be compensated for their contributions. The News/Media Alliance ("N/MA"), for example, has entered into a licensing framework with AI company ProRata through which N/MA's thousands of news and media affiliates can opt to license their content for generative AI uses.[58] ProRata calculates a proportional share of revenue for participating publishers based on the model's use of their content to generate output.[59]

This is not a picture of "market failure," as Meta would have us believe. The market has responded. A determination that Meta's unlicensed appropriation of millions of copyrighted works to build and operate Llama was a fair use could fatally undermine critical current and future licensing opportunities for the publishers of books, journals and other texts whose works are being exploited.[60]

---

[56] *Campbell*, 510 U.S. at 590 (internal quotations omitted).

[57] Katie Paul & Anna Tong, *Inside Big Tech's Underground Race to Buy AI Training Data*, Reuters (Apr. 5, 2024), https://www.reuters.com/technology/inside-big-techs-underground-race-buy-ai-training-data-2024-04-05/.

[58] Sam Quigley, *News/Media Alliance Announces AI Licensing Partnership with ProRata*, News/Media Alliance (Mar. 26, 2025), https://www.newsmediaalliance.org/prorata-licensing-partnership/.

[59] Charlotte Tobitt, *FT, Atlantic, Axel Springer and Fortune Get Behind AI Start-up's Per-use Compensation Plan*, Press Gazette (Aug. 7, 2024) https://pressgazette.co.uk/news/prorata-ai-publisher-deals-financial-times-axel-springer-fortune-atlantic/.

[60] In emphasizing the market for AI training data—the current focus of this litigation—AAP does not mean to suggest that other markets for publishers have not been negatively impacted by Meta's appropriation and inclusion of their works in Llama.

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

Contrary to Meta's efforts to minimize the importance of the AI training market,[61] that Meta does not view that market as a "normal" or "traditional" one for publishers[62] does not negate its salience in weighing the fourth fair use factor. In the pivotal case *American Geophysical Union v. Texaco Inc.*,[63] for example, which addressed then-new photocopying technology, the Second Circuit recognized that unlicensed photocopying of journal articles by Texaco caused market harm because, even though the market was new, publishers had made licenses available for that use.[64] To accept Meta's argument that there can be no market harm unless a licensing market is "traditional" would be to treat the exploitation of copyrighted works by any new technology as undeserving of compensation. Needless to say, that is not the law. In assessing fair use, courts consider whether the market in question is one "that creators of original works would in general develop or license others to develop."[65] It is only logical that copyright owners will continue to pursue licensing agreements with AI companies in a booming market with high demand for their works.

Finally, Meta seeks to dismiss Plaintiffs' claim of market harm as "circular," complaining that Plaintiffs are asserting a right to license a use that Meta insists is transformative.[66] But it is Meta that is relying on circular reasoning: that characterizing a use as transformative eliminates any need to investigate its impact on a market or potential market for the works at issue.

C. A Finding of Fair Use Based on the Exploitation of Pirated Work Contravenes Long-Established Copyright Policy

Meta and its supporting *amici*, the Electronic Frontier Foundation (EFF) and four professors of law, urge this Court to ignore Meta's choice to bypass lawful access to Plaintiffs' works and populate its LLM with pirated material instead in assessing whether Meta's use was

[61] *See* Meta Br. at 30 (calling the AI training market "a potential market" that is "not likely to develop.").
[62] *Id.* at 10, 25-26.
[63] 60 F.3d 913, 930 (2d Cir. 1994).
[64] *Id.* at 930-31.
[65] *Campbell*, 510 U.S. at 592; *accord Seuss*, 983 F.3d at 460 (same); *Hachette Book Grp, Inc.*, 115 F.4th at 192 (legally cognizable markets include "likely to be developed markets").
[66] *See* Meta Br. at 26-27; *see also* Intell. Prop. L. Professors' Br. at 8-11.

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

fair.[67] In addition to evincing bad faith,[68] Meta's conduct is all the more troubling because it contravenes the framework Congress established in 1998 by adding the DMCA to the Copyright Act.[69] With the DMCA amendments, Congress sought to establish a robust digital marketplace by ensuring appropriate safeguards for works made available online, including copyright owners' ability to rely on DRM protections in distributing electronic copies of their works. As explained in the accompanying Senate Report:

> Due to the ease with which digital works can be copied and distributed worldwide virtually instantaneously, copyright owners will hesitate to make their works readily available on the Internet without reasonable assurance that they will be protected against massive piracy. [The DMCA] …. will facilitate making available quickly and conveniently via the Internet the movies, music, software, and literary works that are the fruit of American creative genius.[70]

In weighing Meta's claim of fair use, this Court can and should consider Meta's conduct in relation to the objectives of the Copyright Act and Constitution's directive to protect the exclusive rights of authors in their works.[71] In addition to avoiding the inconvenience and expense of licensing and compensating copyright owners for the commercial use of their content, Meta opted to evade technological protections that are essential to a functioning online marketplace for copyrighted works. This is manifestly at odds with the mandate of Congress in adopting the DMCA. A finding of fair use in this case would not only undermine the public interest in a workable copyright regime, but encourage and reward theft twice over.

---

[67] Notably, while Meta's *amici* urge the Court to ignore Meta's conduct in considering Meta's fair use claim, neither of their briefs expressly advocates for an overall finding of fair use. *See generally* Intell. Prop. L. Professors' Br.; EEF Br..

[68] Plaintiffs' Br. at 8-11.

[69] DMCA, Pub. L. No. 105-304, 112 Stat. 2860 (1998).

[70] S. Rep. No. 105-190, at 8 (1998).

[71] U.S. Const. art. I, § 8, cl. 8 ("The Congress shall have Power … To promote the Progress of Science … by securing for limited Times to Authors … the exclusive Right to their respective Writings."); *see also Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 558 (1985) ("The economic philosophy behind the clause empowering Congress to grant patents and copyrights is the conviction that encouragement of individual effort by personal gain is the best way to advance public welfare through the talents of authors and inventors.") (internal quotes omitted).

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600

CONCLUSION

AAP respectfully submits that this Court should deny Meta's claim of fair use and grant Plaintiffs' motion for partial summary judgment.

DATED: April 11, 2025

Respectfully submitted,

FRANKFURT KURNIT KLEIN + SELZ PC

By: */s/ Jacqueline C. Charlesworth*

Jacqueline C. Charlesworth
Ruby A. Strassman
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
Telephone: (310) 579-9600
Facsimile: (310) 579-9650

Kristen G. Niven*
28 Liberty Street, 35th Floor
New York, NY 10005
Telephone: (212) 826-5566
Facsimile: (424) 372-2947

Attorneys for Amicus Curiae Association of American Publishers

* Not admitted in California or the U.S. District Court for the Northern District of California

Frankfurt Kurnit Klein + Selz PC
2029 Century Park East, Suite 2500N
Los Angeles, California 90067
P (310) 579-9600